冯杰 主编 / 冯吴亮 孙咏萍 帅晓红 副主编

# 普通物理专题研究教程

（第2版）

清华大学出版社

冯杰 主编

冯吴亮 孙咏萍 帅晓红 副主编

# 普通物理专题研究教程

## （第2版）

清華大学出版社

北京

## 内 容 简 介

本书是“物理·课程与教学论暨学科教学(物理)”硕士学位研究生教学计划中学位必修课程“普通物理专题研究”的教材。基于物理学理论和研究方法,主要讨论普通物理学五大学科的基本概念的本质含义、基本规律的相互联系、基本原理的普遍意义、五大学科的基本结构、物理学的研究方法以及现代物理学的量子理论、相对论的物理意义、熵的统计解释等一些新概念、新理论和新实验现象的最基础的知识。立足于对物理·课程与教学论硕士研究生、物理教育和科学教育的本科高年级学生进行物理学理论的基础知识及其研究方法的整合、强化、研讨和提高。

本书可以作为“物理·课程与教学论暨学科教学(物理)”硕士学位研究生和物理教育本科的专业学位课程的教材,也可以供物理教育专业博士研究生、中学物理教师和从事物理教育研究的专业人员参考。

**版权所有,侵权必究。举报:010-62782989,beiqinquan@tup.tsinghua.edu.cn。**

**图书在版编目(CIP)数据**

普通物理专题研究教程/冯杰主编. —2 版. —北京:清华大学出版社,2020.7
ISBN 978-7-302-55944-3

Ⅰ. ①普… Ⅱ. ①冯… Ⅲ. ①普通物理学—高等学校—教材 Ⅳ. ①O4

中国版本图书馆 CIP 数据核字(2020)第 120411 号

**责任编辑:** 朱红莲
**封面设计:** 常雪影
**责任校对:** 刘玉霞
**责任印制:** 刘海龙

**出版发行:** 清华大学出版社
**网　　址:** http://www.tup.com.cn, http://www.wqbook.com
**地　　址:** 北京清华大学学研大厦 A 座　　**邮　　编:** 100084
**社 总 机:** 010-62770175　　**邮　　购:** 010-62786544
**投稿与读者服务:** 010-62776969, c-service@tup.tsinghua.edu.cn
**质量反馈:** 010-62772015, zhiliang@tup.tsinghua.edu.cn
**印 装 者:** 三河市铭诚印务有限公司
**经　　销:** 全国新华书店
**开　　本:** 185mm×260mm　　**印　　张:** 30.5　　**字　　数:** 740 千字
**版　　次:** 2011 年 4 月第 1 版　2020 年 9 月第 2 版　　**印　　次:** 2020 年 9 月第1 次印刷
**定　　价:** 95.50 元

产品编号:086770-01

# 序

“普通物理专题研究”是“物理·课程与教学论暨学科教学(物理)”硕士研究生的学位必修课程,这是该学科培养方案规定的。同时该课程也是物理教育本科和科学教育本科的专业选修课程。所以,该课程的教学目标是使读者进一步加深对物理学原理、体系和研究方法的理解和升华,总体上进一步把握物理学的学科内容,强化物理学理论结构、理论知识和物理学素养。

物理学科课程是培养物理学人才的基石,没有物理学科知识的积淀,则很难造就高层次的物理学人才。对于中学物理教师而言,其今后无论承担中学物理教学还是开展中学物理教育教学研究,都必须立足于物理学的学科基础,否则,就是无源之水、无本之木。这也是物理学科教育必须遵循的规律。近年来,就学科教育领域来看,无论是大学本科还是高端的研究生阶段,都出现了一些令人担忧的现象。其一是作为物理学科的“物理·课程与教学论暨学科教学(物理)”的硕士研究生阶段,由于入学考试的导向等因素影响,生源的物理学科基础非常薄弱;其二是目前我国高等师范院校的物理学(师范)本科生的物理学科基础课和理论课的教学目标达成度比较低。这只是问题的一个方面,另一方面,学科教育的学科基础被大大地弱化了,甚至有脱离学科的不正常现象,这虽然引起了教育界和学科教育专家的注意,但是,直到目前并没有得到有效的改观。

冯杰教授等再版的《普通物理专题研究教程》在回归学科本源、改变“物理·课程与教学论硕士暨学科教学(物理)”硕士研究生和本科生的物理学科教育目标的达成度方面进行了比较长时间的探索和教学实践。其一,该新教程避开了繁难的数学分析和理论物理抽象的逻辑演绎,定位于普通物理学层次,立足于物理学的基本理论,使得物理学基础薄弱的读者能够比较容易读懂。其二,该新教程严谨地界定了对应课程的教学目标不是物理教学研究。既是普通物理专题中的物理学方法论,也是物理学研究所涉及的科学方法,不是中学生学习的认知方法和物理重难点的教学方法。其三,该新教程的一个重要目标是解决研究生和本科生在学习普通物理学理论知识过程中的疑难杂症,力争在普通物理层面上做比较深入、细致和全面的研讨和提高,旨在全面提升其物理学的专业素养。其四,该教程以专题讲座、论坛交流和研究报告等形式对普通物理的逻辑结构进行了重新审视,既展示了物理学的精深理论内涵也彰显了物理学的理论思维价值和巨大的教育功能。

冯杰教授多年从事“普通物理专题研究”课程的教学,具体融入本书中具有以下特点:

1. 浓缩的知识结构:该教程基于清晰的教学目标,注重普通物理知识的结构化。全书首先将物理学归纳为“实物与场”“粒子与波”“时间与空间”“运动与静止”的四对概念范畴,深入浅出系统化地浓缩了普通物理的基本结构,使读者从一个更高的角度,对原有普通物理知识体系有一个整体认识,把握各知识模块之间的相互关系。

2. 研究式的研学方式：该教程各专题尝试运用逻辑学和哲学的相关原理，注重物理学的方法论研究，运用分析、类比、理想化、概括抽象和数学方法等科学思维的方法，引导读者对普通物理学的基本知识、基本原理和基本方法进行研讨和思考，以培养学生理解和体会物理学科学方法的特点。同时每专题(每次课)都有若干与教学内容相关的中心思考题作为课后练习，让学生课后利用课程资源和通过各种方式进行思考、研讨，起到“温故而知新”的效果。

3. 评价形式灵活实效：该教程每章以专题形式呈现知识结构，以讲座或专题报告的形式展开每次课的教的形式，以讨论和资料呈现的方式作为每次课的学的主要形式。教学过程中实时记录学生运用学习资源的有效度、讨论参与度和疑难杂症的消除程度。成绩评定以普通物理学的力学、热学、电磁学、光学和原子物理学中的典型知识点和常见的疑难杂症为主要考查内容。考试题型是论述、讨论、问答和辨析，没有计算题。

4. 注重物理学理论的发展性：该教程在注重物理学理论相对完整和系统性、基本概念和基本原理的基础性的同时，积极发挥教材的教学实用功能和间接传承功能，积极吸收物理学理论研究新进展的新概念、新理论、新实验和新技术原理；积极吸收大学物理学教学研究的新成果，同时考虑到当前中学物理新课程改革的需要，使得该新教程处于具有相对完整和相对开放，又有一定前瞻性的动态系统之中。

目前，虽然有物理学科教育同仁们一直在努力进行大学物理学的教育研究，但是，国内的“普通物理专题研究”仍然处于“初级阶段”，鲜有适合“物理·课程与教学论暨学科教学(物理)”硕士研究生的学位必修课程和物理教育本科和科学教育本科的专业选修课程的教科书。该教程的撰写是一次新的尝试，希望全国同行有更多同类的著述问世，为繁荣我国的物理学科教育做出更大的贡献。

胡炳元

2019 年 10 月于华东师范大学

# 目 录

# 第1章

# 物理学的基本概念范畴

概念是反映思维对象及其特有属性的思维形态。思维对象是指一切能够被人类认识或思考的客观事物。在人类广袤的研究空间里，基本概念具有不同的层次。简要来说，基本概念中涵盖最广泛的是辩证逻辑思维，包括唯物辩证法范畴（物质与意识、运动与静止、时间与空间、对立与统一、质变与量变、肯定与否定、本质与现象、内容与形式、必然与偶然、原因与结果、可能与现实）、认识论范畴（认识与实践、感性与理性、真理与谬误、绝对真理与相对真理）、逻辑学范畴（思维形式、思维规律、思维方法），是具体科学总范畴。

物理学的总范畴包含了所有与物理学研究对象有关的基本概念，小到基本粒子大到天体，从实物物质到场物质，从宏观机械运动到微观量子运动，从引力到核力，从时间到空间等所有有关的物理学研究对象均在其中，由此提炼出物质存在的基本形态——实物和场、物质的根本属性——运动的相关性质描写、物质运动形式的基本描述——粒子和波以及物质存在的基本形式——物理学的时空观等基本概念。最后是物理学中各具体学科理论系统的概念范畴，如力学中的力、质量、惯性、碰撞、参考系等，热学中的温度、内能、功、热量等，电磁学中的电场强度、电势差、位移电流、涡旋电场等，光学中的干涉、衍射、偏振、双折射等，原子物理学中的定态、能级、衰变等。

本章从普通物理学的整体理论体系出发，基于物理学理论的进展，从学习和讨论的角度，将物理学理论的概念体系分析为四对概念范畴，即实物与场、粒子与波、介质波与概率波以及时间与空间，以下各节做较为详细的讨论。

## 1.1 物理学的概念范畴综述

辩证逻辑思维基本范畴是反映各个客观事物具有的一般规定性的思维形式。其三种思维形式是概念、判断和推理。在辩证逻辑里，概念是辩证思维的基本逻辑形式，判断和推理则是概念的内在矛盾的展开，因而，概念是逻辑中最主要、最本质的东西。物理概念是物理抽象思维的最基本形式。

### 1.1.1 概念与物理概念及其表述

概念是人们在认识客观世界的过程中形成的。人们运用感觉器官去观察自然现象，通过感觉、知觉和表象的过程，对有关事物进行比较、分析、综合、抽象和概括等思维活动，区分出一般与个别、本质与现象，抽出一般的、本质的东西形成概念，抛弃了那些非本质的属性。因此，概念是反映客观事物本质属性的一种最基本的思维形式。

**1. 概念**

概念(idea/notion/concept)是人类在认识某一类事物的过程中，从感性认识上升到理性认识，把所感知事物的共同本质特点抽象出来加以概括的思维形式。

作为人类所认知的思维体系中最基本的构筑单位，概念可以是大众公认的，也可以是个人认知特有的一部分。表达概念的语言形式是词或词组。

概念都有内涵和外延。每一个概念都有其确定的内涵及外延。概念的内涵是此概念所包括的一切对象的本质属性，概念的外延就是适合此概念的对象范围。

**2. 物理概念**

概念作为思维的细胞，科学认识首先通过概念来概括和总结。在物理学中，物理概念是反映物质及其物理属性的一种思维形式。例如，我们经常看到人提重物、马拉车、书压桌面等现象。各种现象千差万别，把这些我们所感知的现象，经过比较、分析、综合等思维活动，认识到它们的表现形式虽不同，但是如果丢开这些具体形象，忽略一些非本质的区别，可以发现每种现象至少都包含两个物体，而且一个物体正在对另一个物体施加作用。我们把提、拉、压等动作抽象概括为作用，人、马、书等物体抽象概括为施加作用的物体，重物、车、桌子等物体抽象概括为受到作用的物体，由此得出“施加作用的物体作用于受到作用的物体”。进一步分析，人向上提重物，重物必向下拉人；马向前拉车，车必向后拉马；书向下压桌子，桌子必向上支撑着书。故而说明：物体间的作用是相互的。于是，形成了力的概念：力是物体间的相互作用。又如在惯性概念中，抛弃了“维持运动需要力”这一非本质现象，揭示了物质本身具有保持自己原来运动速度不变的性质。

物理学的研究中，首先建立概念，然后经过判断和推理，建立物理学中的定律和定理，进一步做出新的假设和预言，通过实验检验后，组成完整的物理学理论体系。

每一个物理概念都有其确定的内涵及外延。概念的内涵是此概念所包括的一切对象的本质属性。概念的外延就是适合此概念的对象范围。比如，力的概念的内涵是物体间的相互作用，包括了全部力所共有的本质属性；而力的概念的外延包括了重力、弹力、摩擦力、静电力、洛伦兹力等这些对象的全体。

**3. 物理概念的表述**

物理概念的表述往往通过“定义”“分类”“比较”来体现。“比较”是为明确概念进行的准备工作，“分类”是明确概念的辅助逻辑方法，“定义”是明确概念基本的逻辑方法。

(1) 概念的逻辑学含义

对一个概念下定义时，常用一种 “属概念”加“种差”的方法。属概念是外延较广而内涵较少的概念，当加上足以同其他概念相区别的属性(即“种差”)时，被定义的概念就表述出来了。例如，用(临近)属加种差定义法给“标量”下定义——只用数值就能完全确定的物理量称为标量。这里“物理量”是“标量”的临近属，“只用数值就能完全确定”是“标量”相对于其他“物理量”的种差。所谓“属”和“种差”是逻辑学里的术语，按照分类的方法，事物大类包含几个小类，大类叫“属”，小类叫“种”，上一级种变成下一级属，“种差”即种之间的差别。相对于种，同一级属称为“临近属”，如弹力和摩擦力。

概念从定义的逻辑顺序上可分为基本概念和导出概念。

对一个概念下定义要遵循三条基本原则：第一，定义概念的外延和被定义概念的外延必须相等。例如，“摩擦力是阻碍物体间相对运动的力”的定义概念的外延太宽了。又如“理想气体内能”的定义代替“气体内能”的定义就太窄了。第二，定义不应循环，即“定义项中不能直接或间接包含被定义项”。例如，“质量是物体所含物质的多少”，“能量是物体做功的本领，功是能量变化的量度”，其定义中包含用已知的去定义未知的同义循环，因而不是恰当的定义。第三，定义要全面清楚，不能用比喻。例如，“时间就是生命”，“电容是导体容纳电荷

的能力”,“电压是形成电流的原因,正如水压是形成水流的原因一样”,都不能算作定义。

(2) 概念的比较

比较是弄清不同概念的内涵和外延上的同一性与差异性的初步逻辑方法。根据对象的差异,包括为了抽取物理现象的本质而进行的物理现象要素的比较;也包括相似的物理概念之间的对比。例如,重力与压力、质量与重量、速度与加速度、振动与波动、电势与电势能、干涉与衍射。

(3) 概念的分类

分类是把一个概念的外延进行划分的逻辑方法。例如,经典力学中根据相互作用的物体是否直接接触,可以对力的概念作如下简单分类:

- 力(根据接触与否)
  - 非接触力
  - 接触力(根据力的方向)
    - 摩擦力(切向)
    - 弹力(法向)

### 1.1.2　物理量的定义

物理量是定量化的物理概念,因此,物理概念的定义有其独具的特点,即“完备性”,也就是说,物理概念的定义必须由定性定义和定量定义两部分构成。物理量就是物理概念的定量化描述。

**1. 基本概念物理量的定义方法**

基本概念物理量的定义是基于实验的人为规定,可以不遵循形式逻辑法则。从“完备性”考虑,基本物理量的定义有定性定义和定量定义。

定性定义,是人为规定物质及其运动的某一本质属性。

定量定义,是操作性定义,人为规定单位标准,有时还需人为给定数值的定量计算式。要注意的是,基本物理量是理论系统的基本概念,与国际单位制中的 7 个基本量不同。国际单位制中有 7 个基本量(本书将专题讨论),但全部物理学中“基本概念物理量”却较多。

**2. 导出概念物理量的定义方法**

导出概念物理量的定性定义是由基本概念物理量按上述的形式逻辑法则定义的物理量,反映物质及其运动的某一方面的本质特征。

导出概念物理量的定量定义一般由公式描述,其定量公式通过实验规律确定,反映了导出概念物理量与已知物理量的逻辑关系。

在定义某一导出物理量时,需要注意以下几个问题:

其一,逻辑简单性原则,不要人为杜撰导出概念物理量。我们之所以这样规定物理量的定义式,是考虑了其内在性与规律一致性,以及整个理论系统有可能存在统一而简洁的表达形式。爱因斯坦认为:理论体系的逻辑简洁性与客观世界的客观规律是一致的,而且是需要我们用实践证明的。正由于这种原因,我们总乐于采用使理论具有逻辑简单性的物理量的定义,使理论过于复杂化的定义而引出的物理量,是多余且不必要的。所以,“逻辑简单性”可以作为定义物理量的一个原则。

其二,物理量定量式的选择。这一选择原则上是任意的。所谓定量定义只要能从公式中指出该量的大小、方向即可。通常采用直观上与定性定义一致的定量式,这样容易认识该概念的物理实质,但又要注意该式定义的普遍性原则。例如,对一段导线的电阻的定义,有下面两个定义可供选择,分别为

$$R=\frac{U}{I} \tag{1-1-1}$$

$$R=\rho\frac{L}{S} \tag{1-1-2}$$

其中式(1-1-2)反映“$R$ 是由材料本身决定的电学性质”,较直观,但式(1-1-2)只对某种导线成立(柱形的均匀导体),有局限性。从式(1-1-1)也能讨论出 $R$ 只由材料决定的性质$\left(\text{即不管 } U \text{ 多大,比值}\frac{U}{I}\text{总不变,说明 } R \text{ 只由材料本身决定}\right)$,但初学者感到不直观。可是式(1-1-1)却有它的优点,即普遍性,无论何种形状的导线均可由此式测出 $R$ 值。所以,从理论上讲,式(1-1-1)作为 $R$ 的定义更恰当。

其三,定量式中要能够正确反映该物理量的方向性性质:既是标量,还是矢量,或是张量。

国际单位制的 7 个基本量均是标量,称为基本标量。

基本矢量只有一个,即位矢 $\boldsymbol{r}$。其他物理量的标量、矢量和张量性质均可以由定义或定律结合基本矢量而导出。

### 1.1.3 建立物理概念的基本思维方法

我们知道,物理概念是物理抽象思维的基本形式,但在形成物理概念的过程中,还要依赖于形象思维和直觉思维,依赖于各种思维方法的综合运用。不同的物理概念引入和建立的方法不尽相同。

**1. 运用抽象概括的方法**

在物理学理论中,物质可以有不同的研究层面或侧重点,抽象概括的运用也有不同的具体形式。

(1) 分析概括一类事物的物理共同属性和本质特性,如平动、转动、振动、力等。

(2) 抽象出物质、运动的某一属性,得到表征物质运动某种性质的物理量,如用比值定义法得到的密度、速度、加速度、比热容、电阻、电容、电场强度、磁感应强度等。这些物理量和定义式中的其他各量无关。

(3) 用理想化的方法进行科学抽象,包括创建理想模型,如质点、刚体、理想气体、绝对黑体、纯电阻等。

设立理想实验过程,如自由落体、简谐振动、准静态过程、光的直线传播等。

总之,运用抽象概括的方法建立物理概念的思维程序是:举例——分析——概括——抽象——下定义(定性:文字形式)——公式(定量:数学形式)。

**2. 运用理论推导的方法**

(1) 对有关概念进行概念组合。相关物理量的概念可以组成一个有机的整体,它们之间存在密切的联系,一个物理概念往往是一些物理概念的发展,又是另一些概念的基础。

如质量与速度组合成动量的概念:$m\boldsymbol{v}$;力和时间组合成冲量的概念:$\boldsymbol{F}t$;位置矢量与力组成力矩的概念:$\boldsymbol{M}=\boldsymbol{r}\times\boldsymbol{F}$;角动量:$\boldsymbol{J}=\boldsymbol{r}\times m\boldsymbol{v}$;电通量:$\Phi_E=\iint\boldsymbol{E}\cdot \mathrm{d}\boldsymbol{s}$;等等。

(2) 从物理规律演绎物理概念。物理概念与规律有着不可分割的联系,在许多物理规律(特别是利用实验归纳得出的)表达式中常常存在比例系数,这些比例系数可分为两类:一类是普适恒量,对不同的物质是同一个值,如库仑定律中的 $k$,万有引力定律中的 $G$ 等;另一类则因物质的不同而不同,它反映了物质的某种属性,因而是一个物理量,如摩擦系数

$\mu$，电阻率 $\rho$，自感系数 $L$ 等。除此之外，由于研究的需要，也演绎出了一些辅助的物理概念，如电位移矢量、热力学态函数等。

**3. 运用类比推理方法**

可以使用类比方法进行推理，借助已有的物理概念，来建立新的物理概念。

如由机械波类比光波、无线电波、地震波、德布罗意波(物质波)等。

由引力做功与路径无关或环路积分为零，说明引力是保守力，可以引入取决于系统相对位置的引力势能。利用引力势能概念的建立过程，进一步延伸来分析，弹性力、静电力、分子力做功都与路径无关，因此可以相应地引入弹性势能、静电势能、分子势能等概念。

由库仑定律和万有引力定律的类比，可以在引力场中引入引力场强和引力线。

由波粒二象性、波函数、概率密度等概念，发挥丰富的想象力，把概率密度的分布称为电子云，这样建立了电子云的概念。

**4. 运用等效探究的方法**

有些物理概念涉及的具体内容是变化的，为了研究方便，往往可以运用等效探究的方法。

(1) 将某一物理概念等效为另一相关概念，简化研究步骤，从而引入新的概念。如变速运动可等效为匀速运动或匀加速运动，因此需要引入平均速度或平均加速度。

(2) 直接引入新的物理概念，对某一物理概念的探究直接等效为对这一新的概念的探究。如电路中，电压、电流往往是变化的，为了研究方便，就直接引入了电压有效值、电流有效值等概念。

### 1.1.4　物理学的概念与原理体系

在物理学的众多概念之中，最基本的概念是 7 个基本量，其次是基本的重要物理概念，比如，力学的质量、动量、量子、力、功等，由此构成物理学的定律、定理，从而形成物理学的原理体系。可以做如下概括。

**1. 物理学原理体系的 4 大基本主线**

经过人们长期的实验和理论研究，已经基本形成贯穿物理学发展历程的 4 大主线：①能量守恒(质能关系)；②热力学第二定律；③时间和空间的相对性；④波粒二象性。

本书后续各章节将做详细讨论。

**2. 7 个重要思想**

根据人们在研究自然科学过程中逐步形成的物理认知，可以将经典物理学和近代物理学中的核心思想做历史性的如下概括：

(1) 地球不是宇宙的中心——哥白尼学说；

(2) 宇宙是按照一定规律运动的力学体系——牛顿物理学；

(3) 能量推动着这个系统——能量概念；

(4) 系统运动朝着一个确定的方向——熵与概率；

(5) 事实是相对的，但定律是绝对的——相对论；

(6) 你不能预测或知道所有事物——量子理论和因果限制；

(7) 基本事物不会变化——能量守恒与对称性。

**3. 8 个重要的基本核心概念**

综合力、热、电、光、原子(量子)物理学 5 个学科的概念体系，可以归纳出如下 8 个重要的基本核心概念：①力；②场；③能；④熵；⑤物质的结构性；⑥时间的相对性；⑦粒子；

⑧波。

**4. 4对概念范畴**

物理学发展到相对论阶段和量子理论阶段,人们逐步将经典物理学概念体系进行了升华和更新,无论从物理学理论逻辑关系角度,还是从学习、认知的角度,或者从哲学的高度,可以将整个物理学中的核心思想概括为4对概念范畴,目前国内外的大学物理学体系基本上是按照这4对概念范畴编排的:①实物与场;②粒子与波;③运动与静止;④时间与空间。

以下几节从物理学的4对概念范畴出发,展开讨论物理学基本原理的结构和方法。

## 1.2 实物与场

早在19世纪初叶,法拉第(Michael Faraday,1791—1867)首先提出了场的概念;能量守恒定律是19世纪科学的三大发现之一;但是,20世纪以前,实物和质量、场和能量还被人为地混淆在一起。在这个时期,人们认为实物是由单个的粒子构成的,是由具有一定质量(原子量和分子量)的原子和分子构成的,实物具有间断的结构。在实物结构中不容许任何连续性,场只具有连续性和波动性。实物和场在性质和结构上是绝对对立的,二者之间既没有任何统一,也根本不容许它们相互转化。到19世纪末,关于实物和场的概念,其片面性的谬误开始暴露出来。20世纪初,狭义相对论的质能关系,第一次把实物和质量、场和能量统一在一个简单的方程里,彻底打破了人们认为场和实物之间不能相互转化的物理观念。物质的两种物理形态由原先的分离走向统一的认识过程,是从光学领域开始的。光的波粒二象性和德布罗意物质波的实验事实迫使人们不得不普遍地承认,作为物质两种基本形态的实物和场以及它们结构的共性应当成为物质结构的基础。任何物质(不论是实物还是场)都具有质量,都是间断性和连续性的统一,连续性的波动具有粒子性,实物粒子具有波动性。就这样,人们对实物和场的认识由二者的人为分离达到二者的辩证统一。现在,物质的两种物理形态——实物和场——在结构上既具有粒子性(间断的),又具有波动性(连续的);在性质上既具有质量,又具有能量;而且,二者在一定的条件下可以相互转化。

爱因斯坦(Albert Einstein,1879—1955)说:“实物和场,与其说是定性的区别,不如说是定量的区别。”因为物理学的定性结论是反映物理本质特性的,而物理学的定量只是具体测量方法的计算。

### 1.2.1 实物

**1. 实物的概念和特征**

现代物理学中,实物被定义为“具有静止质量的微观粒子以及由它们所组成的物体”。须注意,实物中包括正实物(正粒子组成)和反实物(反粒子组成)。

实物存在的基本形态至今发现的约有10种。在日常条件下,5态是:

(1) 固态,严格应称为“结晶态”,具有确定的体积、形状、熔点、凝固点等;

(2) 液态,在一定条件下,具有确定的沸点和凝固点,具有一定的体积、没有形状等;

(3) 气态;

(4) 非晶态,例如普通玻璃、橡胶、天然树脂、沥青和高分子塑料等;

(5) 液晶态,具有确定的体积、形状,没有确定的熔点、凝固点等。

在特殊条件下还发现另外5种物态:

(6) 等离子态，超高温下形成的具有正、负等量电荷的一种类似“气态”物质，例如太阳等炽热恒星内部的状况；

(7) 超固态，超高压下的原子被“压碎”，原子核紧密排列形成的，例如恒星后期“白矮星”内部的状况；

(8) 中子态，原子核进一步被“压碎”，中子紧密排列，例如，恒星后期“中子星”内部的状况；

(9) 超导态，在超低温下，物体的电阻趋于零的现象，例如，首次发现水银在温度降到4.173K即约−269℃时，出现电阻为零的现象；

(10) 超流态，物体内部的内黏滞力(摩擦力)突然消失的现象，如液态氦在温度降到2.17K时突然失去黏滞力，出现超流动性。

**2. 实物的一个基本单元层次——基本粒子**

众所周知，物体通常由原子组成，原子又由原子核和核外电子组成。原子核也有内部结构，是由核子——质子和中子构成。这样到了1932年，人们认为所有物体都是由电子、中子、质子和光子组成。在20世纪40年代末期，人们在宇宙线实验中发现了一类所谓的奇异粒子，如K介子、Λ超子、Σ超子、Ξ超子等。这些粒子之所以称为奇异的，是由于它们的奇异量子数$S$不等于零。之后在20世纪50年代末到60年代初，由于高能加速器的建成以及探测技术的迅速发展，除了发现两类中微子$\nu_e$和$\nu_\mu$外，还发现有200多种平均寿命只有$10^{-23}$s左右的粒子，如Δ(1232)、Λ(1520)等。括号中数字是粒子静止质量的能量(单位MeV)，这些粒子统称为共振粒子。从1974年以来，人们又发现了质量为质子质量3倍多而寿命却比共振质子长约$10^3$倍的新介子J/φ(3097)和φ(3685)，还发现了其他的一些重介子如$D^{\pm}$(1870)、γ(9460)等，这些粒子统称为新粒子。这些粒子之所以称为新粒子，是由于组成它们的夸克是u、d、s 3种之外的其他夸克粒子。以上简单介绍的粒子都统称为基本粒子(有关基本粒子的详细内容在下一节中讲述)。

**3. 基本粒子的分类**

对上述的基本粒子，如果按照其间的相互作用种类的不同来进行分类，则可分为如下三类：

(1) 光子：只参与电磁相互作用的粒子。

(2) 轻子：不参与强相互作用，但参与弱相互作用和电磁相互作用的粒子。

(3) 强子：参与强、弱和电磁相互作用的粒子。

基本粒子的绝大多数为强子和轻子。在强子中，自旋为整数的粒子称为介子，自旋为半整数的粒子称为重子。进一步研究发现，强子都是有结构的。1964年，盖尔曼提出所有强子都是由三种更为基本的粒子所构成的夸克模型，这些更基本粒子称为夸克。夸克模型的主要内容有：

(1) 所有强子都由更为基本的夸克粒子构成，每一种夸克具有一种“味”。

(2) 所有介子都由一个夸克(q)和一个反夸克($\bar{q}$)组成。

(3) 所有重子都由三个夸克组成。

该模型在理论上解释了众多的实验事实，但是大量的实验都没找到自由夸克。而对于光子和轻子，至今实验上没有发现它们有内部结构的事实。

**4. 描述一个基本粒子的性质所必需的几个物理量**

(1) 质量：以电子的质量作为一个单位，将其他粒子的质量表示为电子质量的倍数，或

按照狭义相对论的质能关系 $E=mc^2$ 将质量 $m$ 用相应的能量 $E$ 来表示。

(2) 电荷：基本粒子的电荷或与电子相同，或与质子相同，分别以一、+表示。理论研究表明，夸克具有 1/3 e 的电荷量。

(3) 寿命：有些基本粒子能长久地单独存在，称这些粒子为稳定的；有些基本粒子，单独存在一段时间后，就会衰变为其他粒子，它们单独存在的这段时间，称为基本粒子的寿命。

(4) 自旋：每一个基本粒子都有的内禀属性。

(5) 衰变产物：表示一种粒子转化为其他粒子时的产物。

### 1.2.2 场

**1. 场的概念和特征**

1821 年法拉第在关于载流导线绕磁极转动的研究中认识到：磁力是圆形力，圆形力是简单的，且能用于电磁现象的解释。他还认为载流导线周围必定存在着某种“张力”状态，这种张力是通过媒介传递的近距作用，这里已初步包含了“力线”和“场”概念的初步假设。后来的实验工作使法拉第坚信，电和磁的作用不是没有媒介地从一个物体传到另一个物体的“超距作用”。他设想，在磁体、载流导体、带电体的周围空间存在着某种由磁和电产生的像以太那样的连续介质，起着传递磁力和电力的媒介作用，这实际上是“场”概念的萌芽。1845 年他第一次使用了“磁场”这个词，两年后他又单独使用“场”这个词，这是物理学中第一次提出的作为近距作用的“场”的概念。

随着科学的发展和人们认识水平的提高，人们对“场”的概念的认识在逐步深化。今天人们认识的场是物质的另一种形态，电场、磁场、引力场等都是场的一种，具有如下特征：

(1) 没有静止质量，但运动质量不为零。

(2) 具有能量、动量、动量矩等物理量。

(3) 具有波粒二象性。

(4) 服从质能关系和动量及动量矩守恒定律。

(5) 在同一空间中可以同时叠加多种不同的场。

(6) 可以与实物互相渗透、互相影响。

(7) 不能充当参考系。

**2. 场的基本单元——量子场**

(1) 量子场的基本理论

量子化的场就是量子场。量子场作为物质的基本形态，它具有多种运动状态，例如，各种激发态和基态。

各种激发态的量子场，代表在各种运动状态下的基本粒子系统。量子场激发态的出现，代表实粒子的产生，量子场激发态的消失，代表实粒子的消失。实粒子的消失只是表明量子场的激发态的消失，而不是量子场这种客体本身消失了，作为物质客体的量子场当然还存在着，不过此时量子场是处于能量最低的运动状态，也就是处于基态。处于基态的量子场，由于实粒子的数目为零，因而称为零场，也就是真空。由此可见，场与粒子是高度统一的，并且每一种粒子都联系着一种场，或者说，每种粒子都相应于该粒子所从属的量子场的元激发，例如，光子就对应于电磁场的元激发。

由于粒子是从属于量子场的，因此两者的运动规律具有相连性。在量子场论中，量子场和粒子的运动规律是随机决定的。表示运动规律的场方程，同时也反映了能够在该量子体系中发生的这样或那样量子过程的可能性，由于这些不确定性，我们不能把量子场论中的基

本粒子说成通常意义的粒子。另外,规范场论认为所有场的运动规律具有定域规范对称性,并且获得巨大的成功,于是人们便把规范对称性这个约束提高到原理的地位,称为规范场对称原理,来作为规范场论的一个公理。

(2) 量子场论的发展和超统一理论

麦克斯韦(James Clerk Maxwell,1831—1879)在19世纪60年代使电力和磁力统一于一个单一的电磁场理论。20世纪上半叶爱因斯坦曾经花费了他后半生的时间和精力,把大自然的四种基本力(引力、电磁力、弱作用力和强作用力)统一起来。但是或许由于没有考虑量子效应,爱因斯坦的努力没有成功。

1961年到1968年间,美国物理学家戴维·格罗斯(S. L. Glashow,1932— )、斯蒂芬·温伯格(Steven Weinberg,1933— )和阿布杜斯·萨拉姆(Abdus Salam,1926— )在杨振宁与非阿贝尔规范场(Yang-Mills 场)的基础上独立地建立了弱电统一理论,第一次成功地改造了电磁力和弱作用力的数学表达式,使这两种力被结合进一个统一的数学表达式之中(1979年,他们在统一力的研究方面的工作获得了诺贝尔奖),并预言了 $W^+$,$W^-$,$Z^0$ 粒子。弱电统一理论就是把电磁相互作用和弱相互作用看作一种相互作用,在相互作用过程中,如果交换的是 $W^+$,$W^-$ 和 $Z^0$ 粒子,就是弱相互作用过程;如果交换的是光子,就是电磁相互作用过程。此时,W,Z 玻色子还没有被观察到,但人们相信,这只是迟早的事。果然,1983年,欧洲核子研究中心(CERN) 发现了它们。弱电统一理论得到了充分的肯定。

在统一力的理论中,出现了更加抽象的对称群。人们早就知道,麦克斯韦优美的电磁理论之所以有力量、之所以优美,在很大程度上要归功于该理论的数学描述中所显示出来的平衡和对称。这种平衡被称作规范对称,是一种抽象的平衡。

物理学家们受到统一规范理论成功的鼓舞,把注意力转向了另一种核力——夸克间的色动力,不久之后,就提出了色规范理论,接着,有人便试图将弱力和色动力统一到一个"大统一理论"(grand unified theory,GUT)中去,办法是使用更大的规范对称将所有的其他对称包容在一个规范对称之中。目前,估计 GUT 的成就还为时尚早,但至少它所作的一个预测——经过无限长的时间之后,质子可能会很不稳定并自发地衰变——现在正有人进行检验。

随着大统一理论的提出,现代量子场论的研究也在快速发展,正在不断完善这4种相互作用的力的动力学理论,并得以进一步探讨它们统一的物理机制。现在,物理学家越来越倾向认为,建立一个超统一的理论,来完成对4种相互作用力的统一。那么,超统一理论也应当能完全描述一切夸克和轻子。有了超统一理论,我们就能够用一个方程式,用一种宇宙的总公式把大自然的一切秘密都写下来。

图1-2-1是场的统一理论发展的图表表示。

### 1.2.3 实物与场的比较

#### 1. 实物与场的相同之处

我们已经介绍了实物和场的一个基本单元层次——基本粒子和量子场。作为物质的两种基本形态,实物与场有许多相同之处:

(1) 实物有不同的形式与状态,场也有不同的形式与状态。

(2) 实物有微粒结构,场也有微粒结构,如电磁场的微粒是光子。

(3) 场和实物都具有波粒二象性,只是在特定情况下,它们的两种性质显现程度不同。标志波动性的频率 $\nu$、波长 $\lambda$ 和标志微粒性的能量 $E$、动量 $p$ 通过普朗克常量 $h$ 定量地联系

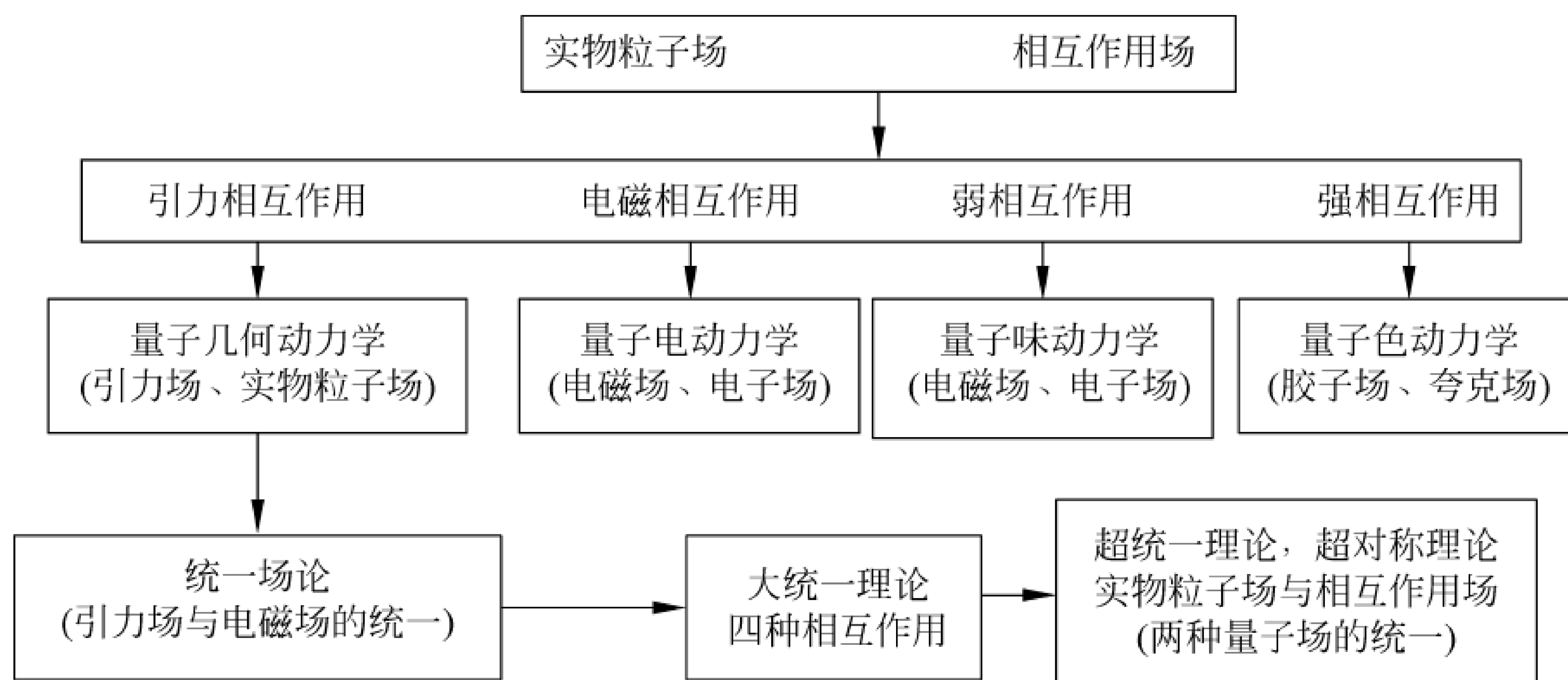



图 1-2-1 场的统一理论结构图

起来，即关系式 $E=h\nu$ 和 $p=h/\lambda$。

(4) 场和实物一样，不仅有能量、动量等，也有质量。

(5) 场和实物都遵守各种守恒定律及质能关系，因此场的产生必定是以物质的某种形态消失为前提。

**2. 实物与场的区别**

实物与场虽然有共性，但无论从形态上还是从相互作用上，也存在很大的区别，具体可归纳如下：

(1) 实物的质量和能量高度集中在有限空间内，场的质量和能量一般是弥散在无限空间内。

(2) 实物的粒子有静质量和动质量之分，场粒子只有一个不变的运动质量 $m=h\nu/c^2$，而无静质量。

(3) 就波粒二象性而言，实物粒子的粒子性较强，场的波动性较强。对于实物粒子，根据 $\lambda=h/p=h/mv=\dfrac{h\sqrt{1-v^2/c^2}}{m_0 v}$，通常情况下，$v\ll c$，所以 $\lambda=\dfrac{h}{m_0 v}$，$\lambda$ 数量级较小，在一般实验条件下是显现不出来的；而对于场，如光波，最能体验到的一般是它的频率、波长。它的波动性实验总是与它的宏观现象相联系，并反映它的宏观物理性质；而其粒子性实验总是与电子等微观粒子或物质的微观结构相联系，并反映作为微观粒子而存在的光子的物理性质，这种性质较宏观性质不易被观察到。

(4) 场空间是兼容的，即一个空间可以同时容纳几种场，如引力场、电磁场可以在同一个空间内存在，而且通常它们确实就存在于同一个空间内，而实物所占空间不能同时容纳其他实物。

(5) 场和实物可以同时占有一个空间。如带电粒子在电场里要受到电场力的作用，粒子与场处于同一个空间。

(6) 实物之间的作用是整体行为，是有“后效”的，如实物之间碰撞后能量、动量发生改变，而场之间的作用是局域的，是没有“后效”的，干涉现象就是例证，两束相干光发生干涉以后，性质无任何变化，我们无法从这些性质判断出它们是否相遇过。

(7) 实物与场在功率的传递上区别很大。

由此可见，场与实物的相同本质使得在可预见的未来将会发展出两者统一的理论，但同

时也由于场与实物之间有如此多的区别,致力于此理论发展的人们仍面对着重重的困难。

### 1.2.4 实物与场的统一

以狭义、广义相对论和量子理论的诞生为标志,经典物理学已经进入到近代物理学阶段,这是一次重大的理论上的变革,一些经典物理学的基本概念和基本原理受到了冲击,而代之以更高层次的理论上的认识。在现代量子场论之前的近代物理学的探索中,物理学理论(主要是广义相对论和量子力学)的看法是:物质有两种基本存在形态——实物和场。这时,实物被定义为"具有静止质量的微观粒子以及由它们所组成的物体"。须注意,实物中包括正实物(正粒子组成)和反实物(反粒子组成)。而场则被定义为"传递实物间相互作用的物质形态"。相应于不同实物间的相互作用,至今物理学家认识到有4种场:引力场、电磁场、强相互作用场、弱相互作用场。从宏观看来,实物与场在存在形态上的主要区别是质量(注意不是仅指静止质量)或能量在空间中相对的集中,有较为确定的接口,占据有限的空间,有"不可入性",是一种物质"集聚"的形态。而场表现为物质在空间中的连续分布且具有可叠加性,描述场的物理量满足波动方程,可给出各处场量随时间的变化规律,这就是场波动性的特征。从经典观念来看,实物具有粒子性,场具有波动性,二者从概念上是不能兼容的。

但物理学的发展,却逐步在走向这两个概念的统一。从微观上看来,实物与其相应的相互作用场间事实上不能截然分开,因为实物与场的明确接口无法确定,例如,质子和中子总为一个属于其结构的虚$\pi$介子外壳所包围,与强相互作用场不能截然分开。因此,爱因斯坦说:"实物与场之间的区别,与其说是定性的问题,倒不如说是定量的问题。""实物便是能量密度特别大的地方,场便是能量密度小的地方。""把实物和场看作是彼此完全不同性质的两种东西是毫无意义的。"微观物理学的实验进一步说明,实物与相互作用场之间紧密联系。从存在的角度看,任何实物必与相应的相互作用场同时存在(不但实物外,实物内粒子间也有相互作用场,虽然相互作用场有时可以脱离实物"场源"而独立存在,例如,即使光源已经消失,光源发出的光仍可在空中传播);从相互转化角度看,正反实物粒子与相互作用场间在一定条件下可以相互转化(例如正负电子对光子的相互转化等)。

**1. 量子理论对实物与场概念统一的初步研究**

量子理论首先在物质的粒子性和波动性的统一上进行了研究。从经典物理学看来,粒子性和波动性对同一物质客体来说是不能兼容的。但是量子理论的发展却一再证明物质可以同时具备粒子性和波动性,经典理论中两个不兼容的概念在一种新的意义上可以统一,从而进一步消除了实物与场在概念上的鸿沟,为两种存在形态的统一做出了贡献。爱因斯坦基于普朗克"能量子假设"提出的"光量子假设",说明光(即一种电磁场)不仅具有波动性,而且具有粒子性;之后,德布罗意的"物质波假设",说明实物粒子(静止质量不为0的微观粒子)不仅具有粒子性,还具有波动性。进而推广为一切微观粒子(无论静止质量为0与否)都同样具有粒子性和波动性。这为实物与场在概念上的统一铺平了道路。

作为微观物理学的基础理论,量子力学还有它的不足:

(1) 主要侧重实物粒子的波粒二象性,甚至没有把光当作粒子处理;

(2) 在粒子的基础上谈其波动性——概率波——波粒二象性地位不平等;

(3) 实物粒子的波动性——概率波——不是真正物理波场;

(4) 不能描述粒子的产生与消失——微观粒子的转化。

**2. 量子场论对实物与场的统一的贡献**

量子场论是在狭义相对论和量子力学基础上发展起来的。量子场论的研究方向重新回到了爱因斯坦设想的主题：将实物对应于场，场是最基本的物质存在形态。它不仅坚持了而且从更高的角度强调了“相互作用场”的波粒二象性；也提出了“实物粒子场”的新概念，以在存在形态上实现实物与场在理论上的统一，并且从相互转化的角度也说明了“实物粒子场”与“相互作用场”间的统一。

量子场论关于物质存在基本形态的观点是：

(1) 相比实物，场是物质存在更基本的形态，粒子是场量子的激发态；

(2) 对于实物粒子场而言，场量子是实物粒子，如电子场、中子场等；

(3) 对于相互作用场而言，场量子是媒介粒子，如引力场、电磁场等；

(4) 真空态是各种量子场均处于基态的状态；

(5) 实物粒子之间的相互作用表现为场之间的相互作用；

(6) 任何量子场都是真实的物理波场，可以用波动方程描述；

(7) 用算符描述其运动规律。

量子场论中，粒子与场在概念上是统一的，所以量子场之间的相互作用就是相应粒子间的作用。一个量子场的激发即反映了粒子的产生；一个量子场的去激发即反映了粒子的消失；一些量子场的去激发伴随另一些量子场的激发则反映了粒子之间发生了转化。所以，量子场的相互作用是粒子产生、湮灭、相互转化等过程的动力学根源和基本机制，较好地解决了用场的运动规律去描写粒子运动规律的要求。

## 1.3 粒子与波

粒子与波都是人们对物质运动形式的阶段性认知。我们知道，物理学理论的建立和发展经历了经典物理学和近代物理学两大阶段，人类对物理学概念范畴的认识经历了逐步深化的发展过程。比如，对实物与场及物质的波粒二象性等的认识，经历了漫长而又曲折的发展历程，如今，对实物粒子结构的认识，对实物粒子的波动性认识仍然处于不断的探索和深化当中。

### 1.3.1 粒子概述

电子、质子、中子是比原子核更深一层次上的物质存在形式，在这些粒子间存在着四种基本作用。20 世纪 60 年代以后人们又认识到这些基本粒子具有内部结构。研究基本粒子和夸克等的基本性质、内部结构以及相互作用和相互转化规律的学科称为“粒子物理学”。

从古代开始东西方就在探询宇宙中万物的组成，如《道德经》中“天地万物生于有，有生于无”和“道生一，一生二，二生三，三生万物”。从老子书中所表达的“道”的意思来看，大体上是指宇宙的本源，而“一”是指原始的混沌状态，“二”是指阴阳，“三”是指阴阳相激，产生气。另外《圣经》的创世纪中也有这样的文字：大地混沌，还没有成形。深渊一片黑暗；上帝的灵运行在水面上。上帝命令：“要有光，光就出现。”这是第一天。两千多年前，古希腊唯物主义哲学家，原子唯物论学说的创始人之一德谟克利特(Democritus，前 460—前 370)就提出了原子论；中国也有“金木水火土”五行学说；1800 年，英国化学家、物理学家、近代化学之父道尔顿(John Dalton，1766—1844)提出了近代原子论，认为一种元素代表一种原子；元素周期表中有 109 种元素，其中最后 17 种由人工生成。

1896 年，法国物理学家贝克勒尔(Antoine Henri Becquerel，1852—1908)发现铀的放射

性,标志着原子核物理学家研究的开始;1897年汤姆孙(Joseph John Thomson,1856—1940)发现电子;1898年法国科学家居里夫妇(皮埃尔·居里,Pierre Curie,1859—1906;居里夫人,1867—1934)发现放射性更强的元素钋和镭;接着在1903年英籍新西兰物理学家卢瑟福(Ernest Rutherford,1871—1937)提出放射性元素蜕变理论;1919年,卢瑟福又首次实现了人工核反应,并发现了质子;1932年英国物理学家查德威克(James Chadwick,1891—1974)发现了中子,同年苏联物理学家伊万年科(Д. Д. Иванечко,1904—1994)和德国物理学家海森伯(Heisenberg,Werner Karl,1901—1976)各自独立提出原子核是由质子和中子组成的假说,日臻完善了原子由质子、中子、电子组成的三子模型,如图1-3-1所示。

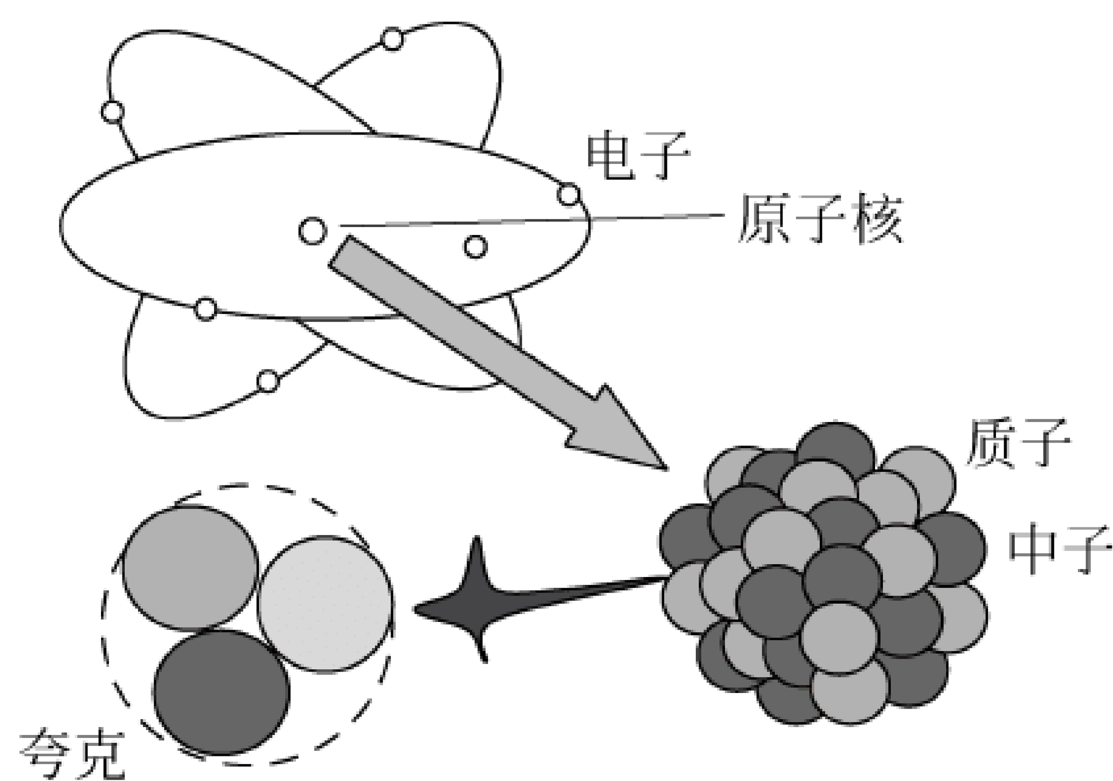


图1-3-1　三子模型

现在人们一般都认为:物质是由分子构成的,分子是由原子构成的,原子包括核外电子和原子核,原子核是由质子、中子等基本粒子组成的,而基本粒子则是由比基本粒子更基本的亚粒子组成的。亚粒子也就是人们常说的“夸克”。

我们可以把这些粒子分成3类:费米子、玻色子、希格斯粒子。

**1. 费米子**

费米子共有48种,自旋是$\hbar$的半整数倍的基本粒子称为费米子,包含轻子和夸克。典范的轻子就是电子,电子与另一种叫作中微子(严格说应是电子中微子)的轻子相伴生。当电子参与放射衰变这类过程时,总有中微子产生。由于一些目前还无人知晓的原因,这一基本图像已经复制了两次,产生了三“代”轻子。

除电子本身外,还有比较重的叫作$\mu$介子,它除了比电子重207倍外,完全像是电子;还有一种甚至更重的粒子叫作$\tau$粒子,它的质量接近质子的两倍。这两种粒子各有其自己的中微子,所以轻子族有六种(三对)粒子。虽然$\mu$介子和$\tau$粒子都能在粒子加速器中用能量制造或从宇宙线产生,但它们很快衰变,转化成电子或中微子。

另外与电子及它的中微子相对应的两种夸克(它们的名字很怪,叫作“上”夸克和“下”夸克),由它们就能解释质子和中子的结构。一个质子由通过强力维持在一起的两个上夸克和一个下夸克构成,而一个中子由通过强力维持在一起的两个下夸克和一个上夸克构成。

强力本身可视为胶子的交换,而胶子本身又由夸克对组成,因而又称为介子。正如轻子族复制了三代,夸克族也如此。虽然只需要两种夸克来解释质子和中子的本质,但复制的两代夸克却一代比一代重,其中一代叫作奇夸克和粲夸克,最重的一代叫作底夸克和顶夸克(表1-3-1)。

**表1-3-1　六种夸克特性**

| 夸克 | 电荷/e | 自旋 | 同位旋 | 奇异数 | 重子数 |
|---|---|---|---|---|---|
| u(上) | 2/3 | 1/2 | 1/2 | 0 | 1/3 |
| d(下) | −1/3 | 1/2 | 1/2 | 0 | 1/3 |
| s(奇) | −1/3 | 1/2 | 0 | −1 | 1/3 |
| c(粲) | 2/3 | 1/2 | 0 | 0 | 1/3 |
| b(底) | −1/3 | 1/2 | 0 | 0 | 1/3 |
| t(顶) | 2/3 | 1/2 | 0 | 0 | 1/3 |

和重轻子一样，这些夸克能够在高能实验中产生(因而大爆炸时必定大量存在过)，但迅速衰变成它们的较轻对应物。虽然不可能分离出单个夸克，但粒子加速器实验已经提供了夸克族所有这六个成员存在的直接证据；最后一种(顶)夸克是芝加哥费密实验室的科学家于1994年找到的。

对夸克的质量和其他性质的研究表明，不可能有更多代的夸克，只能有三族夸克和三族轻子。幸而标准大爆炸模型也认为不可能存在多于三代的粒子；不然的话，极早期宇宙中额外中微子造成的压力应该驱动宇宙过快地膨胀，从而使留存下来的氦含量与极老恒星的观测结果不符。但是，除了大爆炸的最早片刻之外，第二和第三代粒子在宇宙的演化或其内容物的行为中基本不起作用。

我们在宇宙中看到的每样东西都能用两种夸克(上和下)和两种轻子(电子和电子中微子)加以说明；确实，由于单个的夸克不能独立存在，我们看到的每样东西的行为，仍然能够用1932年就已经知道的电子、中子和质子再加上电子中微子，以及四种基本力，相当准确地予以近似说明。

为了满足泡利不相容原理，说明夸克有另外一个自由度，即每一种夸克又必须有三种不同的状态，取名为颜色：红、绿、蓝，三基色混合出白色。强子都是白色的。

不同种类的夸克称为味道，夸克有六味：u，d，s，c，b，t；每味有红、绿、蓝三色(R，G，B)。与阴阳相符，费米子中的每个粒子都是配成对的，即都有正、反粒子。正、反粒子的一部分性质完全相同，另一部分性质完全相反，例如电子和正电子是一对正、反粒子，它们的质量和自旋完全相同但电荷等值反号，磁矩方向相反。6个轻子有6个反粒子共12个，六味三色18个夸克有18个反粒子共36个，所以费米子为12+36=48种。

**2. 玻色子**

玻色子共有13种，自旋是$\hbar$的整数倍的基本粒子称为玻色子。如果把费米子比作“砖头”，那么玻色子就是“水泥”，因为它们是四种基本相互作用的媒介粒子。

按照目前的理论，宇宙间的一切物质的相互作用归结为表1-3-2中4种。

强相互作用是一种原子核结合在一起的力，它的作用范围为$10^{-15}$m以内，故称短程力；作用时间极短，约为$10^{-23}$s。这种相互作用强度大，比电磁作用强得多，故称强作用。强相互作用只存在于重子和介子之间。

弱相互作用范围在$10^{-17}$m以内，通常也称短程力。弱相互作用的强度比强相互作用弱得多，作用时间慢得多。参与弱相互作用的粒子包括一切强子和轻子。

一切具有电荷或磁矩的全部粒子包括强子和轻子都参与电磁相互作用。电磁相互作用通过电磁场交换光子来实现。一些基本粒子通过电磁相互作用发生衰变。电磁相互作用的强度比强相互作用弱，电磁力的大小与物体间距离的平方成反比。

$$F=k\frac{Q_1Q_2}{r^2},\quad k=9\times10^9\,\mathrm{N\cdot m^2\cdot C^{-2}} \tag{1-3-1}$$

所有粒子都参与万有引力相互作用，其强度比弱作用还小，与物体间距离的平方成反比。

$$F=G\frac{m_1m_2}{r^2},\quad G=6.67\times10^{-11}\,\mathrm{N\cdot m^2\cdot kg^{-2}} \tag{1-3-2}$$

胶子是传递强相互作用的媒介粒子，共有8种。光子是传递电磁作用的媒介粒子，质量为零，不带电，自旋为1。中间玻色子共3个($W^+$，$W^-$，$Z^0$)，是传递弱相互作用的粒子。传递

引力的粒子称为引力子，自旋应为 2，静止质量和电荷为零，以光速运动，尚未被实验证实。

表 1-3-2　4 种相互作用

| 力 | 强作用 | 电磁作用 | 弱作用 | 引力作用 |
| --- | --- | --- | --- | --- |
| 范围 | 核内 | 原子 | 核内 | 天体 |
| 强度 | 1 | $10^{-2}$ | $10^{-9}$ | $10^{-38}$ |
| 媒介 | 胶子 | 光子 | $W^{\pm}$，$Z^{0}$ | 引力子 |
| 自旋 | 1 | 1 | 1 | 2 |
| 个数 | 8 | 1 | 3 | 1 |

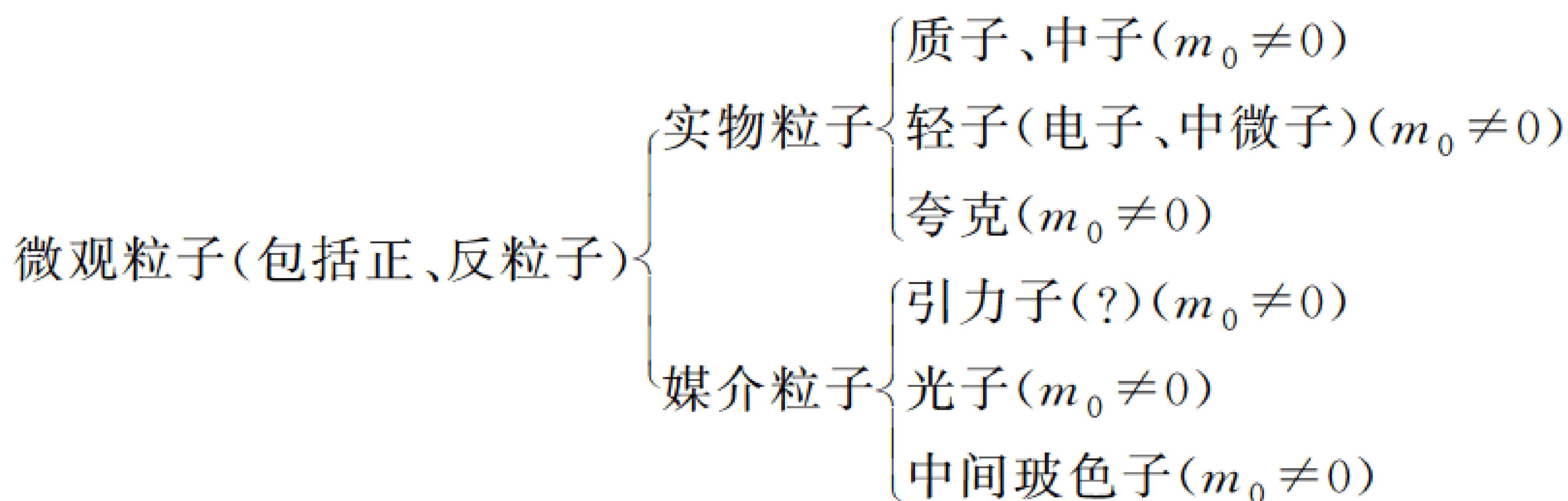
目前的状况是弱相互作用和电磁相互作用已经找到了运动状态变化之间的联系，统一了。但它们与后两种力之间有没有运动状态变化之间的内在联系，人们尚一无所知，在某种意义上说，寻求更大的统一实际上是寻求运动状态变化之间的内在联系。4 种基本力能否最终统一，还需要一代一代的年轻物理学家继续努力。

**3. 希格斯粒子**

希格斯粒子是美国物理学家希格斯(P. W. Higgs)在 1964 年提出的。理论上对于它的所有相互作用性质和运动行为都有精确的描绘和预言，但对于其质量却没有给出任何预言。现在已经在实验中发现这种粒子，由于实验的发现，希格斯荣获了 2013 年诺贝尔物理学奖。从实验结果分析，希格斯粒子的自旋为零，质量应大于 $58.4\mathrm{GeV}/c^2$。

综上所述费米子有 48 种，玻色子有 13 种，外加希格斯粒子共有 62 种。物质世界就是由这 62 种粒子构成的。现在还没有任何实验迹象表明这些粒子有内部结构。

以下是目前物理学在理论上对微观粒子在最小层次上的分类：以 $m_0$ 表示静止质量，问号“?”表示还未经实验证实。




## 1.3.2　波概述

**1. 经典机械波的分解与本质**

经典机械波的运动学方程(又称波函数)

$$y = A\cos\omega(t - x/u) \tag{1-3-3}$$

$$y = A\cos\frac{2\pi}{\lambda}(x - ut) \tag{1-3-3$'$}$$

式(1-3-3)或式(1-3-3′)可以分解为两部分：振动与平动。即波函数是由 $y$ 轴上的振动方程

$$y = A\cos\omega t \quad 或 \quad y = A\cos\omega(t - t') \tag{1-3-4}$$

与 $x$ 轴上的平动方程

$$x = ut' \quad 或 \quad t' = x/u \tag{1-3-5}$$

复合而成的。由此不难看出“振动与平动的矢量和”是波动的起因和本质。

**2. 粒子波的产生及其广义波动方程**

对于连续介质,振动加平动可以产生波动。而对于不连续的物质(非介质)——粒子将会怎样?对于以平动方程(1-3-5)运动的平动粒子,若同时有一振动施于其上,且振动方程为式(1-3-4),则作用于该粒子的平动与振动的矢量和,就复合成了可以用波动方程(1-3-3)或式(1-3-3′)表述的波动运动的粒子——波动粒子或粒子波。其运动痕迹是连续的,像经典波的轨迹。但是它的运动轨迹却随着粒子的定域性而断续,是定域的、显粒子性的粒子的波动。

既然机械波的波动方程式(1-3-3)或(1-3-3′),也是粒子波的波动方程,那么粒子波也可以用复数形式

$$y=y_0\mathrm{e}^{-\mathrm{i}2\pi(\nu t-x/\lambda)} \tag{1-3-6}$$

表示。即式(1-3-3)、式(1-3-3′)或式(1-3-6)称为"广义的波动的运动学方程",而不再是狭义的、仅仅局限于古典连续的机械波的波动运动学方程。式(1-3-6)对新的、不连续的、定域的粒子波也适用。

若对式(1-3-3)取二阶偏导,则可得到粒子波的一维平面波的波动的动力学方程,其一般形式为

$$\frac{\partial^2 y}{\partial x^2}=\frac{1}{u^2}\frac{\partial^2 y}{\partial t^2} \tag{1-3-7}$$

三维空间中传播的平面粒子波的经典波动的动力学方程为

$$\frac{\partial^2 y}{\partial x^2}+\frac{\partial^2 y}{\partial y^2}+\frac{\partial^2 y}{\partial z^2}=\frac{1}{u^2}\frac{\partial^2 y}{\partial t^2} \tag{1-3-8}$$

波动的动力学方程实际是经典粒子物理和波动物理的统一体,是运动学与波动学的统一。波动学是运动学的一部分,是运动学的延伸,即平动与振动的矢量和。对象不同,一个是连续介质,一个是定域的粒子,都可以具有波动性。

**3. 介质波与概率波**

(1) 介质波

经典机械波(简称经典波)是连续介质的扰动,换句话说,没有媒介经典波也就不存在,因此,鉴于这一特点,经典波可以有一个新称呼——媒介波或介质波。

介质波的构成有两个主要因素,一个是物质,即连续介质;另一个是属性,即平动加振动的扰动。介质是物质。波动是"平动+振动"的一种运动形式,是属性,即能量。因此,介质波需要介质(媒质),介质粒子的振动——就是能量的传播,同时介质粒子没有迁移。

(2) 概率波

在经典物理学中我们已经知道,一个被看作质点的宏观物体的运动状态,是用它的位置矢量和动量来描述的。但是,对于微观粒子,由于它具有波动性,根据测不准关系,其位置和动量是不可能同时准确确定的,所以,我们也就不可能仍然用位置、动量以及轨道这样一些经典概念来描述它的运动状态了。微观粒子的运动状态称为量子态,是用波函数 $y(\boldsymbol{r},t)$ 来描述的,这个波函数所反映的微观粒子波动性,就是概率波。概率波无需介质(媒质),波动粒子可以脱离波源独立传播——能量的传播;波的强弱并非是波动粒子的真实行为,只是波动粒子可能出现的时空概率。

以下是物理学对波的分类简表:

两类波
- 介质波　机械波:声波、水波
- 概率波　波粒二象性:光波、物质波(德布罗意波)

### 1.3.3　波粒二象性

**1. 自由粒子的波粒二象性**

所谓自由粒子即与其他粒子的相互作用可以忽略的粒子。显然根据牛顿定律，自由粒子作匀速直线运动。从波动的角度看它对应平面波，若自由粒子的动量为 $p$（设其沿 $x$ 方向，$p=p_x$），能量为 $E$，其频率为 $\nu$，波长为 $\lambda$，则由波粒二象性关系决定：$E=h\nu=\hbar\omega$，$p=\dfrac{h}{\lambda}$（或 $\boldsymbol{p}=\hbar\boldsymbol{k}$）。该自由粒子波函数常表示为

$$\psi=\psi_0\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}(Et-\boldsymbol{p}\cdot\boldsymbol{r})} \tag{1-3-9}$$

或

$$\psi=\psi_0\mathrm{e}^{-\mathrm{i}(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})} \tag{1-3-10}$$

必须注意：用平面波函数来表示自由粒子的波函数只是一种近似。理想的平面波是不存在的。然而平面波方程(1-3-9)或(1-3-10)之所以十分重要，是因为它不仅近似地表示出自由粒子的行为，而且还可以借助傅里叶级数和傅里叶积分通过不同频率的平面波函数叠加构成任意粒子的波函数。克莱因-戈登方程（Klein-Gordon equation）和薛定谔方程(Schrödinger equation)就是在该理论基础上建立起来的。透彻理解平面波函数以及与其有关的自由粒子的波粒二象性，需要借助量子力学的知识。

**2. 有关波粒二象性的新认识**

按照经典物理的概念，依据对原子客体做观测所得到的实验结果的描述，似乎原子客体有时呈现波的性质，有时又呈现粒子性质。例如电子，云雾室照相径迹表现它像个粒子，但晶体衍射实验中它又像波。又如X线，完全类似的晶体衍射表现它像波，但光电效应、康普顿效应等又表现它像个粒子。

按通常理解，粒子是实物的集中形态，波是实物的散开形态，但实物不能同时是粒子又是波，两者明显是矛盾的。一个粒子在某地它就不能同时在另一地，但波（也就是场）就是在一个广延的空间范围中同时发生。

这个矛盾怎样才能解决呢？德布罗意(Louis Victor de Broglie，1892—1987)及量子力学的其他创始人，对于物质粒子波动性的见解，都曾经深受经典概念的影响，把物质波看成是经典概念下的波，粒子只是许多波组合起来的一个波包。波包的大小即粒子的大小，波包的群速度即粒子的运动速度，波包的活动表现出粒子的性质。但这样一幅图画被实验否定了。波包是不同频率的波组成的，不同频率的波在媒质中的速度不同，这样一个波包在媒质中会逐渐扩展而消灭。但实验中观察到的电子不会在媒质中扩展而消灭。再则，波在两媒质的接口上可分为反射和折射两部分，而一粒电子是不可分的。

此外，在电子衍射实验中，电子波打到晶体表面后发生衍射，衍射波将沿不同方向传播开去。如果把一个电子看成三维空间的物质波包，则在空间不同方向观测到的只能是“电子的一部分”，这与实验完全矛盾。实验上测得的（例如计数器或照相底片上记录到的）总是一个一个的电子，各具有一定的质量和电荷等。显然，物质波包的观点夸大了波动性一面，而实际上抹杀了粒子性一面，是片面的。

另一种设想是，粒子是基本的，波只是大量粒子分布密度的变化。波动性是由于大量粒子分布于空间而形成的疏密波。它类似于空气振动出现的纵波，即由于分子密度疏密相间而形成的一种分布。这种看法也与实验矛盾。实际上可以做这样的电子衍射实验，让入射电子流极其微弱，电子几乎是一个一个地通过仪器，但只要时间足够长，底片上仍将出现衍

射花样。这表明电子的波动性并不是许多电子在空间聚集在一起时才有的现象,似乎波动性是各个电子具有的性质。

以上事实表明,运用经典波的概念,是难以将粒子与波统一到客体上去的,然而究竟应该怎样正确理解粒子与波动两重性呢?下面就波动性与粒子性的关系作简要叙述。

波动性与粒子性的争议是历来已久的事,最早是人类对光的粒子性与波动性的争议,牛顿提出粒子学说,惠更斯(Christiaan Huygens,1629—1695)等人提出波动学说,并且被托马斯-杨(Thomas Young,1773—1829)的实验证实;特别是麦克斯韦的电磁波理论,将光归于电磁波中的一部分,对光有了更加清晰的认识。爱因斯坦提出光子的概念,将光的认识统一起来,光具有波、粒二象性。后来,德布罗意发展了波粒二象性,将光的波粒二象性,推广到微粒,提出德布罗意波,又称为物质波。

物质具有二重性,这不合逻辑,到底哪一个性质是主要的,哪一个性质是表现出来的,这是物质的一个本质问题,让我们从物质的基本粒子出发来解决这个问题。有一种观点认为:粒子性是根本的性质,波动性是物质存在时,不断与环境作用光子所表现出的波动性。

所有物质的单独存在没有意义,包括物质的基本粒子——光子的存在本身没有意义,光子只有在运动中,被其他物质不断吸收、发出、与其他物质作用才能表现出自己的质量。物质本身没有静止质量,只有这个物质与环境或其他物质作用光子才能将自己的质量、存在和构成表达出来。其到底是多少,取决于单位时间内物质与环境作用光子的总能量。在单位时间内,作用的光子能量越少,物质对外表现出的物质质量就越少。这可以用一个数学公式来表达:$\Delta E=\sum_{i=1}^{n}h\nu_i$,其中 $\nu$ 是物质吸收和发出光子的最高频率,$\Delta E$ 是在单位时间内物质与环境作用的光子总能量,它越大,$\Delta m=\frac{\Delta E}{c^2}$就越大,我们就感觉到物质质量越多,就是物质的质量大,惯性多;相反也是成立的。

由于物质的存在都要不断地与环境作用光子,没有作用光子,物质的所有性质都表现不出来,这就带来一个问题,作用光子的时候,何处有光子,何时作用,都是不可预测的事,因而产生一个测不准问题,要对物质的行为进行测量,就要用一仪器对这种物质进行作用,即让仪器的光子信息与这种物质的光子信息进行作用,就会不同程度上影响这个物质的行为,这就是量子力学中的测不准原理,其中物质的质量越小,这种影响相对越大,测不准现象越明显。这种测不准原理存在的本质原因就是物质的存在原理,不断地与环境作用光子也是物质存在时的必然现象。测不准现象,是物质波动性的又一个体现。既然测不准原理是由于物质的粒子性在物质存在的过程中不断地与环境作用光子造成的结果,因而光的波动性、物质的波动性、德布罗意波(物质波)等都是由于粒子性引起的,物质存在的过程中不断地与环境作用光子,由于光子的不确定性才表现出波动性。

我们通过一个实例来说明光的波动性是由于粒子性引起的,在某种条件下,波动现象会减弱,而粒子性会加强,甚至波动现象会消失,仅表现出粒子性。在中学物理课本中,有这样一个演示实验,如图 1-3-2 所示,让一束光通过一个小孔,不断缩小小孔的直径,开始的时候,照在屏幕上的光斑的直径不断缩小,但是,当小孔的直径小到一

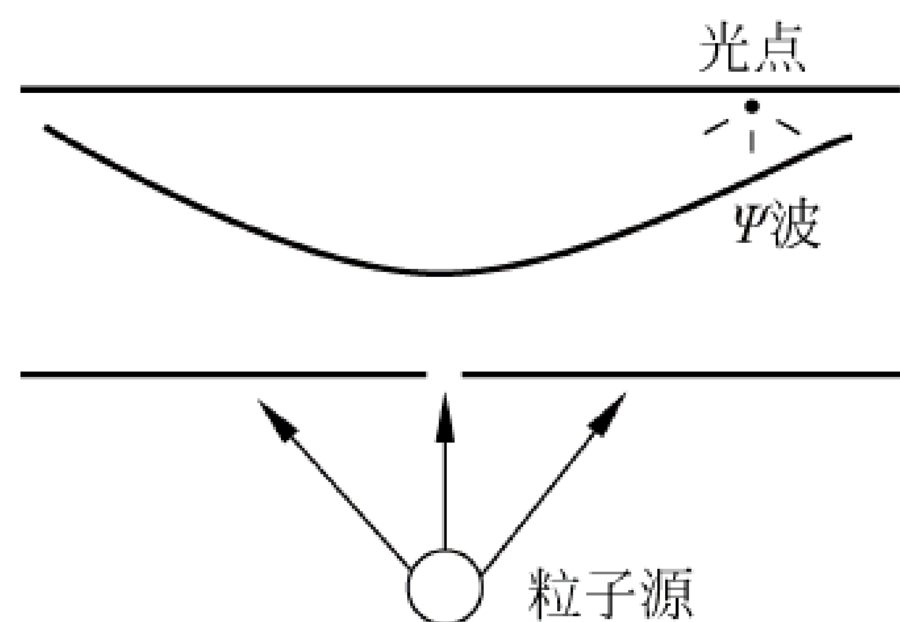


图 1-3-2 小孔衍射实验

定程度，光斑的直径不但不缩小，反而变大，这是物理学中所说的光的衍射现象。还有著名的泊松亮斑，都能很好地说明光发生了明显的衍射现象，就是光的能量传播到阴影区域里去了，这种现象是直接支持了光的波动性，也是托马斯-杨等人反对牛顿的光子粒子学说的直接证据。

在这个实验里，之所以发生了波动，是由于光子在运动过程中，不断地被其他物质吸收和发出，特别是发出过程中，并不能唯一确定是向哪个方向发出的，只能以到达某一个区域的比较多和到达某一个区域的比较少进行比较，所以出现了概率波和波动性这个说法，特别是小孔的直径越小，达到与光子波长接近的时候，被小孔周围物质吸收、作用、发出的可能性越大，走向阴影区域的可能性越大，即物理学中所说的衍射现象越明显。

通过以上分析可知，从粒子与波相对应的角度来看，就光这种物质来说，光子的粒子性是根本的，而光的波动性是光子存在过程中一种表现。同样其他微粒子的粒子性也是绝对的，波动性是相对的，是物质存在过程、表现过程的一种体现。

## 1.4　介质波与概率波

人们对介质波与概率波的认知经历了相当长而且曲折的过程。广义上说，具有时间和空间周期性而且在一定条件下满足叠加原理的物理现象就是波动。即沿波的传播方向经过某一空间距离后会出现同一振动状态（如质点的位移和速度）；在不同介质的接口上能产生反射和折射，遵守反射定律和折射定律。从经典物理学的观点，某一物理量的扰动或振动在空间逐点传递时形成的运动就是波。被传递的物理量扰动或振动有多种形式，因而产生各种形式的波，例如，机械振动的传递构成机械波，电磁场振动的传递构成电磁波（包括光波），等等。但是，从近代物理学的观点，扰动的物理量不同，各种波产生的原因也不同。从波产生的角度来看，介质波与概率波是两种不同类型的波，其产生机制、波函数代表的物理意义等都有本质的不同。

### 1.4.1　介质波

介质波是指某一物理量的扰动或振动在介质中的传播形成的波。经典粒子观认为，粒子多是以直线或圆周曲线运动的，经典机械波是连续介质的扰动。换句话说，没有媒介经典波也就不存在，媒介是经典波的精髓和关键。介质波的构成有两个元素，一个是物质，也就是波赖以传播的介质；另一个是属性，即物理量的扰动，如对于经典机械波，属性乃"平动加振动"。

（1）行波的产生及传播

如图 1-4-1 所示，把一根橡皮绳的一端固定在墙上，用手沿水平方向拉紧，当手猛然向上抖动一次时，就会看到一个突起状的扰动沿绳向另一端传去。在这个例子中，绳就是扰动（波动）传播的介质。绳子可以看成由很多小段的绳子组成，各段绳子之间都有相互作用的弹力联系着，当用手向上抖动绳子左段的第一个质点时，它就带动第二个质点向上运动，第二个又带动第三个，依次下去。当手向下拉动的第一个质点回到原来位置时，它同样也要带动第二个质点回来，而后第三个质点、第四个质点等也将被依次带动回到各自原来的位置。结果，由手抖动引起的扰动就从绳的左端传到了右端。简单的介质波就是这样产生并在介质中传播的。复杂的介质波，产生的机制及传播复杂很多，但传

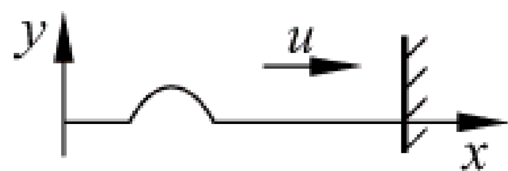


图 1-4-1　橡皮绳产生的波

播的原理跟上面的例子是一样的，都是扰动在介质中的传递。

(2) 波函数

用来描述扰动在介质中传播所形成的波的数学函数式就是波函数。上述的波可用 $y=f\left(t-\frac{x}{u}\right)$ 来表示，它表示绳上的任何一 $x$ 点在任何一个时刻 $t$ 的位移 $y$。

由简谐振动引起并在介质中传播的波动为简谐波。波函数为 $y=A\cos\omega\left(t-\frac{x}{u}\right)$，其中 $A$ 为简谐波的振幅，$\omega$ 为简谐波的角频率，$u$ 为波的传播速度。对波函数，分别对 $t$ 和 $x$ 进行二次求导，再比较可得 $\frac{\partial^2 y}{\partial x^2}=\frac{1}{u^2}\frac{\partial^2 y}{\partial t^2}$，这个微分方程为波动的动力学方程。在这里它虽然是由平面简谐波的波函数求导而成的，但可以从数学上普遍证明它是各种平面波所必须满足的微分方程式，它的解就是平面波波函数的解。它是物理学中最重要的方程之一。它的普遍意义在于：任何物理量，不管是力学量、电学量或其他的量，只要它与时间和坐标的关系满足上述方程式，则这一物理量就按照波的形式传播。

### 1.4.2 概率波

**1. 概率波及其波函数**

(1) 德布罗意的假设及实验依据

法国的物理学家德布罗意认为自然界在很多方面都有明显的对称性，在光的波粒二象性的启发下，他认为如果光具有波粒二象性，那么一般的实物粒子，如电子，也应该具有波粒二象性。他提出这样的问题："整个世纪以来，在辐射理论上，比起波动的研究方法来，是过于忽略了粒子的研究方法；在实物理论上，是否发生了相反的错误呢？是不是我们关于'粒子'的图像想得太多，而过分地忽略了波的图像呢？"德布罗意提出假设——实物粒子也具有波动性，并假设光子的频率和波长分别为 $\nu=\frac{E}{h}=\frac{mc^2}{h}$ 和 $\lambda=\frac{h}{p}=\frac{h}{mv}$，其中 $\lambda$ 称为德布罗意波长。

德布罗意的假设在三年后被戴维森(Davisson, Clinton Joseph, 1881—1958)和革末(L. H. Germer, 1896—1971)的电子散射实验所证实。在实验中，戴维森和革末使一束电子射到镍晶体的特选晶面上，同时用探测器测量沿不同方向散射的电子束的强度，在实验中发现，当入射电子的能量为54eV时，在 $\varphi=50°$ 方向上散射电子束的强度最大。按类似于X射线在晶体表面衍射的分析，得出散射电子数极大的方向应满足条件 $d\sin\varphi=\lambda$。已知镍晶面上原子间距为 $d=2.15\times10^{-10}$ m，从而得出电子波的波长应为 $d\sin\varphi=\lambda=1.65\times10^{-10}$ m。按德布罗意的假设，求出电子波的德布罗意波长为 $\lambda=\frac{h}{m_e v}=1.67\times10^{-10}$ m。

同样，在1927年，汤姆孙做了电子束穿过多晶薄膜的衍射实验，成功地得到了和X射线通过多晶体薄膜后产生的衍射图样极为相似的衍射图像。

另外，戴维森做的电子单缝、双缝、三缝等衍射实验，得出的明暗条纹，更加直接说明了电子具有波动性。

除了用电子做的实验，后来还有用中子、质子等粒子做的实验，都证明了德布罗意的假设及德布罗意公式的正确性。

(2) 概率波的波函数的物理意义

对于德布罗意波的意义是什么，曾存在着不同的讨论。当前得到公认的关于德布罗意

波的实质的解释是玻恩(Max Born,1882—1970)的解释。在玻恩之前,爱因斯坦曾提出电磁场是一种“鬼场”,这种场引导光子的运动,而各处电磁波振幅的平方决定在各处的单位体积内一个光子存在的概率。玻恩发展了爱因斯坦的思想,保留了粒子的微观性,而认为物质波(德布罗意波)描述了粒子在各处被发现的概率。

为了定量地描述微观粒子的状态,量子力学中引入波函数 $\psi$。一般情况下,波函数是空间和时间的函数,即 $\psi=\psi(x,y,z,t)$。波函数本身没有物理意义,而 $|\psi|^2=\psi\psi^*$ 就是粒子的概率密度,即在 $t$ 时刻,在 $(x,y,z)$ 附近单位体积内发现粒子的概率,$\psi$ 因此称为概率波。

**2. 概率波的波动方程**

概率波作为一种类型的波动,应该有对应的波动方程。1925 年,美籍荷兰物理学家、化学家德拜(Peter Joseph William Debye,1884—1966)让他的学生薛定谔(1887—1961,德文:Erwin Schr&ouml dinger,英文通常写作 Erwin Schrödinger,奥地利物理学家。概率波动力学的创始人)作一个关于德布罗意波的学术报告之后,提醒薛定谔:“对于波,应该有一个波动方程。”几个月后,薛定谔经过努力钻研,拿出了一个概率波的波动方程——薛定谔方程,即

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2}+U(x,t)\psi=\mathrm{i}\,\hbar\frac{\partial\psi}{\partial t}\quad(\text{一维薛定谔方程})\tag{1-4-1}$$

其中 $\psi=\psi(x,t)$ 是粒子在势场 $U(x,t)$ 中运动的波函数。利用薛定谔方程就可以解出粒子在任何时刻的波函数。薛定谔方程描述了势场中粒子的状态(波函数)随时间变化的规律,反映了微观粒子的运动规律,它在量子力学中的地位,如同牛顿方程在经典力学中的地位和麦克斯韦方程在电动力学中的地位。

## 1.4.3 概率波与介质波的异同

**1. 两种波函数代表的意义完全不同**

对于介质波,波函数 $y=y(\boldsymbol{r},\boldsymbol{u},t)$ 描述的是受扰的物理量随时间和空间的变化,是定量描述波动过程的数学表达式。例如,行波的波函数是 $y=f\left(t-\frac{x}{u}\right)$,简谐波的波函数是 $y=A\cos\omega\left(t-\frac{x}{u}\right)$。波函数代表了介质中任何一个质点在任何一个时刻的位移的情况。而对于概率波,波函数 $\psi(x,t)$ 本身不代表任何物理量,没有任何物理意义,其绝对值的平方 $|\psi|^2$ 则正比于在时刻 $t$,在空间 $r$ 处找到粒子的概率。

**2. 两种波函数都满足叠加性**

介质波和概率波都具有叠加性,具体表现为空间上的干涉、衍射及驻波。从原理上讲,概率波的干涉现象来源于描述波函数的薛定谔方程是线性的。如果 $\psi_1$ 和 $\psi_2$ 是物理体系的两个可能状态,它们的相干叠加 $\psi=\psi_1+\psi_2$ 也代表物理体系一个可能状态。依据波函数的玻恩解释,其空间表示的模平方 $|\psi|^2=|\psi_1|^2+|\psi_2|^2+\psi_1\psi_2^*+\psi_1^*\psi_2$ 代表了在某空间点发现粒子的概率密度。对一个量子相干叠加态而言,其模平方中包含了不同分量的交叉项 $\psi_1\psi_2^*+\psi_1^*\psi_2$,它们代表了量子干涉,意味着态的量子相干性。仅就数学结构而言,这与介质波或经典光波的干涉是一样的。然而,量子干涉或量子相干性有不同于经典干涉的十分奇妙的基本量子特性。一个典型例子是物质波的双缝干涉实验:实验中,如果测量粒子通过了哪一个缝,干涉条纹便不再存在了。这种由于测量或其他影响导致相干性消失的现象

称为量子退相干。

**3. 两种波的波长、频率、波速**

对于介质波，波长代表的是在介质中两个相邻相同点之间的距离；频率代表的是单位时间内物理扰动的次数；波速代表的是单位时间内物理扰动在介质中传播的距离。即以上各个量都代表了具体的物理意义。而对于概率波，并不是任何实在的物理量在空间的波动，也没有任何实在的物理量在空间传播，波函数中的波长、频率、波速（包括相速和群速）等都包含在复指数之中，其没有直接的物理意义，只是一种表述上的需要。因此，甚至可以出现相速度比光速大的情况。

**4. 概率波具有波粒二象性特征**

概率波（或物质波）可以说是由光的波粒二象性理论而得到的。在光电效应实验之后，受到普朗克关于黑体辐射量子论的启发，爱因斯坦认为光是由光子组成的粒子流，即光具有粒子性；而在19世纪，通过光的干涉、光的衍射等实验，人们已经普遍认识到光是电磁波，并建立了光的电磁理论——麦克斯韦理论。综合起来关于光的本性的全面认识就是：光既具有波动性又有粒子性，即波粒二象性。光既不是经典意义上的"单纯的"波，也不是经典意义上的"单纯的"粒子。在宏观的传播过程，光突出地显示出其波动性；而在微观的相互作用过程，光突出地表现为粒子性。

## 1.5 时间与空间

时间与空间作为物理学中两个基本的概念，从古至今，一直是无数科学家追寻的最高目标；而时空观的不断发展，贯穿在整个物理学的发展历程之中。深入地探讨时间与空间的概念，不仅有助于正确理解这两个物理量，而且能够体验科学的严谨性与科学的美。

### 1.5.1 时间的概念及其单位

时间的意义是如何产生的？我们人类生活在地球上，地球对太阳自转一周的现象称为一日或一天，也可以用年、月来计算。"日出而作，日落而息"。一日、两日、三日……就这样永无休止延续下去的现象，给予我们人类一种感受意识，我们把这个意识叫作"时间"。所以时间意识是以地球为基础的，区分事物运动变化的快慢先后过程。

时间的概念是人类认识、归纳、描述自然的结果，其本意原指四季更替或太阳在黄道上的位置轮回，《说文解字》曰：时，四时也；《管子·山权数》说：时者，所以记岁也。随着认识的不断深入，时间的概念涵盖了一切有形与无形的运动，《孟子·篇叙》注："谓时曰支干五行相孤虚之属也。"可见，时是用来描述一切运动过程的统一属性的，这就是时的内涵。由于古代人们研究的问题基本都是宏观的、粗犷的、慢节奏的，所以只重视了"时"的问题，后来因为研究快速的、瞬时性的对象需要，补充进了"间"的概念，于是，时间便涵盖了运动过程的连续状态和瞬时状态，其内涵得到了最后的丰富和完善，"时间"一词也就最后定型了。

西方哲学家认为时间是伸向过去和将来的无穷远的数轴，而中国古代哲人则认为时间是一个滚滚的"年轮"，靠天干地支排列的轮回，每60年叫一个"甲子"。天干地支是生命不断轮回和消耗能量的对象。在汉语中，"时"指的是一般时间，它表示时间的本质，即运动或过程的持续性。这个以"日"和"土"作为部首和偏旁的汉字（繁体），与太阳的运行和生命的生长有密切联系。它表示或者孕育着生命的本质和本原。时间与阴阳的变化有着密切关系，它们都是以日或月作为偏旁的字。时间就是阴阳交替变化的形式。在我们的传统文化

中，时间表示太阳的运行、自然的过程、生命的过程和生活的过程以及这些过程的延续。在古代人们看来，时间就是这些过程的持续或这些持续的过程或过程(阶段)中的一点。总之，他们只能认识具体的时间，不可能抽象出时间的科学定义，就像他们只能认识事物，不能概括出物质定义来一样。

时间是一个极其抽象的概念，至今看来，对它最科学的“定义”，是由辩证唯物主义给出的哲学“定义”：时间是运动着的物质的存在形式，反映了物质运动过程的持续性(长短)以及它的顺序性(状态先后)。一切物质运动过程都具有的持续性和不可逆性构成了它们的共同属性，这种共同属性就叫作时间。所谓持续性，包括了过程的因果性和不间断性；所谓不可逆性，则指过程能重复但不可能返回过去的性质。时间是客观的，时间的存在和流逝不依人的意志为转移。

时间作为基本概念的物理量，人类人为地规定了其单位。“秒”作为计时的单位，具有一段发展的历史。时间单位秒，最早是利用地球自转运动来规定的。天文工作者把太阳连续两次通过观察者所在子午线上空的时间定为 24 小时，叫作“太阳日”。但是实际上地球自转的周期是不稳定的，太阳日有长有短，因此，国际上就采用一年内各太阳日的平均值作为 24 小时，称为“平太阳日”。1767 年开始使用格林尼治视太阳秒。视太阳秒是指视太阳日的 24×3600 等分之一。1834 年开始使用格林尼治平太阳秒。平太阳秒是指平太阳日的 24×3600 分之一。1960 年开始使用历书秒。历书秒是指 1900 年 1 月 0 日 12 时的平太阳秒。1958 年 1 月 1 日 0 时，国际天文学界就建立了第一个原子时系统。1967 年国际天文学界定义了原子秒。1969 年建立了全球统一的原子时(IAT)，采用了原子秒作为一秒的长度。原子秒是时间的基本国际单位，它的定义是：铯 133 原子基态的两个超精细能级间跃迁对应的辐射的 9192631770 个周期的持续时间。

从时间单位的演变过程可以看到，随着人类认识世界的不断深入与精细，物理概念的单位元越来越精密、越准确。

### 1.5.2　空间的概念及其单位

对空间特性的认识，是古人自然观的重要组成部分。中国古人在探索自然的过程中，形成了丰富多彩的空间观念。

在对空间概念的抽象上，古人有多种说法，常见的是用空间的三维性来定义空间。《管子》书中有《宙合》篇，后人解曰：“四方上下曰合”，“合”即为抽象意义的空间概念。由“四方上下”显然是强调的其三维性。除了用“合”表示空间以外，古人更多地则是以“宇”表示空间。据《文子・自然》篇记载：“老子曰：往古来今谓之宙，四方上下谓之宇。”这就是一例，它也是着眼于空间的三维性的。在历史上，类似说法很多，表明它是中国古代常见的一种空间表示方法。中国古代空间观念的另一特征，是其对空间取向性的重视，认为空间各向异性，方向具有绝对意义。这一特征的作用首先在于它发展出了一套实用角度概念体系，用四维、八干、十二支表示二十四个地平方位角。中国古代缺乏角度概念，这一体系成功地解决了物体分布地平方位的表示问题，因而具有较大的实用价值。而其更重要的意义则在于：这是对空间物理特性的揭示。物质的分布及运动与空间分不开，所谓空间的各向异性，实际上反映了物质在空间中分布及运动的特征，认为它不是混混沌沌、各向同性，而是沿各个方向有不同的表现。因此，我们说这一认识揭示的是空间的物理特性。

辩证唯物主义给出的空间的定义：空间是运动着的物质的存在形式，空间反映了物质存在的地位(位置)及它的广延性。该定义科学地反映了空间概念的本质。

同样的，作为基本概念的物理量，空间长度的单位——米也是人为规定的。“1 米”最初的标准为本初子午线长度的二千万分之一，也就是由南极或北极到赤道距离的一千万分之一。但由于后来发现地球不是正球体，所以“1 米”的定义改为某一合金制的金属尺的标准长度。随着激光的出现，“米”的定义改为以光波的波长倍数作标准。但由于测不准原理的缘故，我们没有可能同时得出长度和速度的精确量度。所以，在 1983 年，国际度量衡组织改以光速为长度标准的定义，将光速定为 299792458 米/秒(准确)。由 1983 年开始，“1 米”被定义为平面电磁波在 1/299792458 秒的持续时间内在真空中传播行程的长度。而在 1960 年所制定的标准，即氪的橙红射线的波长则作为参考标准。

### 1.5.3 物理时空观的进展

物理学时空观与哲学时空观的主要区别是其带有明显的“相对性”。即随着物质分布层次的深入(宏观到微观)或拓展(宏观到宇观)以及运动状态的改变(由低速到高速运动)，物理时空观必定不断进行变革，越来越反映实际，但每一相应的时空观都有其适用范围(局限性)。

时空观的变革可以说是科学理论重大变革的基本标志。物理学研究的发展表明，科学理论的重大变革往往以时空观为突破口，并伴随新时空观的产生。

**1. 物理学的旧时空观**

旧时空观以前的古代时空观是“天圆地方”的“盖天说”——一种宇宙有限论，但盖天说没有摆脱神学的影响，认为宇宙之外是上帝的栖身之地，宇宙是由上帝创造的。

旧时空观的主要观点反映在“地心说”中，到托勒密时期，它已经形成了较完整的宇宙结构学说。以“地心说”为代表的旧时空观的主要观点是：

(1) 承认时间、空间是客观存在的。“地心说”虽然承认“神灵”的存在，但并没有说神灵在支配时空，而是力图从物质本身去寻求运动的原因。

(2) 在空间概念中，正确地认识到“上”“下”观念是相对的。但是，它的宇宙空间却是有限的，是有中心的。

(3) 承认时间的无限性。因为它至少承认了天体圆运动的永恒性，但没有说明宇宙有无起源，即时间有无起点的问题。被神学利用后，强加于“地心说”的时间起点是神创宇宙之时，这并非“地心说”原来的观点。

**2. 物理学的经典时空观——伽利略变换和牛顿的时空观**

对宇宙的认识，哥白尼(Nicolaus Copernicus，1473—1543)用“太阳中心说”代替了“地球中心说”。尽管只是参考系的一个变更(将参考系由地球转向了太阳)，但是该理论显示了以太阳为惯性系描述太阳系内行星运动的优越性，从而展现了宇宙的这一层次——太阳系结构的本来面目，也为牛顿(Isaac Newton，1642—1727)在动力学方面的研究铺平了道路。但是它的缺点是：仍然是宇宙中心说论，并承认了太阳的绝对静止；并限于当时的观测及认识，认为恒星不动。因此，以牛顿为代表的科学家进一步发展了物理学的时空观，创立了经典时空观。

(1) 伽利略变换和牛顿的时空观

伽利略(Galileo Galilei，1564—1642)通过望远镜的观察，进一步说明了太阳系内部结构的统一性，并利用笛卡儿坐标系这一数学工具，给出了关于距离(与空间概念有关)和时间概念的确切的数学形式。与太阳中心相连接的坐标系被称为伽利略坐标系。伽利略提出了伽利略相对性原理：“力学定律在所有惯性系中都相同。”以及对这一原理的数学补充——伽利略变换，即两个惯性系之间的时间和空间坐标的变换式。牛顿在前人大量工作的基础

上，定量地、完整地建立了经典力学的理论体系，为整个经典物理学打下了坚实的基础。按照伽利略变换，牛顿定律完全满足伽利略相对性原理的要求，即三者完全协同一致。因此伽利略变换反映的时空观，集中体现了经典物理学的时空观。

伽利略变换是两个惯性系描写同一物体运动时，时间和空间坐标的变换式，体现了运动描写的相对性。

伽利略变换集中地反映了牛顿的时空观，牛顿的时空观抛弃了旧时空观的“宇宙中心论”，这是很大的进步，其主要观点总结如下：

其一，承认时间、空间客观存在。

其二，时间和空间与物质及其运动无关。

时间坐标系和空间坐标系是完全脱离物质而独立存在的；时间间隔与空间间隔在不同的惯性系中保持不变；即时间、空间观念与物质运动状态无关。

其三，时间和空间彼此无关，各自独立存在。

牛顿体系对应的空间观念是：空间是绝对存在的，并且在各个方向上是无限延伸的、均匀的。具体地说，空间是处处相同的、无差别的、连续的统一体，是由抽象的几何点构成的，物质可用理想化的“质点”替代，并在这个空间中用一个点表示出来。物体的运动表现为点的位置的变化，这样描述的运动是连续的和无限可分的。这样的空间用三维笛卡儿坐标系描述最合适。牛顿体系的时间结构是：时间是一维的、具有单向性，时间是均匀的，同时是绝对的，不因参考系的变换而改变。这种时空结构是平直的，其几何特性在数学上用欧几里得几何描写。

(2) 牛顿经典时空观局限性

**问题之一：“极限速率”的问题。**

由于电子的质量与其电荷相比，比值非常小，所以，它很容易被加速到很高的速率。例如，在真空管的两个电极之间加100V的电压，从阴极飞出的电子(设从静止出发)，将以$6\times10^6$m/s的速率到达阳极：$qU=\frac{1}{2}mv^2$，所以，有$v=\sqrt{\frac{2qU}{m}}=\sqrt{\frac{2\times100\times1.6\times10^{-19}}{9.1\times10^{-31}}}$ m/s$=5.92\times10^6$m/s。

在这种情况下，牛顿定律成立。我们可以把速率的平方表示为与电压的正比关系

$$v^2=\frac{2qU}{m}\propto U \tag{1-5-1}$$

但把电子放在加速电压为几百万伏的情况下，用上述方法对电子速率的平方的计算结果，却与实验测量值产生了很大的偏差。实验采用范德格喇夫起电机，使电子获得1.5MeV的能量，然后注入直线加速器，最终获得15MeV的能量，通过测量电子的飞行时间和路程可以计算出电子的速度。实验数据如表1-5-1所示。

**表1-5-1　计算电子速度实验数据表**

| 动能/MeV | 飞行时间/$10^{-6}$s | 电子速率/$10^8$m/s | 电子速率平方/[$10^{16}$(m/s)$^2$] |
|---|---|---|---|
| 0.5 | 3.23 | 2.60 | 6.8 |
| 1.0 | 3.08 | 2.73 | 7.5 |
| 1.5 | 2.92 | 2.88 | 8.3 |
| 4.5 | 2.84 | 2.96 | 8.8 |
| 15 | 2.80 | ≈3 | ≈9 |

实验结果表明：

其一，被加速电子的速率有个上限，$3\times10^8$ m/s。

其二，电子的动能 $E_k=qU$ 是真实的，但 $v^2$ 达不到预期值。

将方程(1-5-1)表示为 $mv^2=2qU$，则只能是 $m$ 的值在一定的范围内随速率的增大而增大。这显然和牛顿力学的理论相违背。伽利略的速度相加公式 $\boldsymbol{v}_j=\boldsymbol{v}_q+\boldsymbol{v}_x$，是牛顿时空观的数学表述，按照这种表述，我们可以通过一个假想的实验来验证一下，电子在一个静止的参考系中经 0.5MeV 加速电压加速后速率达到 $2.6\times10^8$ m/s，然后进入一个和它保持相对静止的参考系再次被加速，在这个和它保持相对静止的参考系中电子的相对速度为 $v_x=2.6\times10^8$ m/s。按伽利略的速度相加公式，此时电子相对地面的速度应为 $v_j=5.2\times10^8$ m/s，但前面的实验表明，具有这样速率的电子我们无法获得。这表明牛顿力学的规律不能描述高速运动的物体。

**问题之二：质量的变化问题。**

牛顿力学中认为质量有两种：其一，惯性质量，是指物体(平动)惯性大小的量度；其二，引力质量，是指物体产生引力场和受引力场作用能力的物理量。可以证明上述两种质量是相等的。在牛顿力学中无论是惯性质量还是引力质量对于给定物体来说都是不变量。上面提到的电子被加速的实验中已反映出，高速运动的物体的质量随速度的变大而增大，对于高速运动的物体质量不再是恒量。这个现象在牛顿力学中无法得到解释。

**问题之三：高能粒子 μ 子的运动寿命。**

牛顿时空观认为：空间绝对平直，时间绝对均匀，而且空间、时间与物质的运动状态无关。但是，在宇宙射线的研究中发现一种高能粒子 $\mu$ 子，$\mu$ 子是在地球的大气层的上层产生的，静止 $\mu$ 子的寿命只有 $2.1971\times10^{-6}$ s，即使以接近光速的速度运动也只能飞行大约 660m，但在地球表面却能观测到 $\mu$ 子(地球的大气层厚度大约为 $10^4$ m)。这个现象在牛顿力学中也无法得到解释。可能的解释只有两种：首先，在和 $\mu$ 子保持静止的参考系看来 $\mu$ 子的寿命只能是静止寿命，$\mu$ 子之所以能飞到地球表面，是因为在这个参考系看来地球的大气层在运动，因而厚度变薄；其次，在地球的参考系看来，大气层厚度不会改变，$\mu$ 子之所以能飞到地球表面，是因为在这个参考系看来 $\mu$ 子在运动，$\mu$ 子运动的寿命可能比静止寿命长。显然这两种观点在牛顿时空理论中是没有的。

**问题之四：惯性系的局限性。**

在牛顿力学中所有的惯性系都是等价的，在不同的惯性系中所有的力学规律都具有相同的形式(协变性)，但对光及电磁学的规律却不适用。

**问题之五：微观粒子的测不准原理。**

按照牛顿观点，一个粒子的初始位置和初始速度确定之后，其后任意时刻粒子的运动状态都可以确定，这表明粒子的坐标和动量可以同时确定。但在空间尺度较小的微观领域，粒子遵循的却不是这样的规律。它们的行为需要满足测不准原理：

$$\Delta x\Delta y\geqslant\frac{\hbar}{2} \tag{1-5-2}$$

式(1-5-2)表明，微观粒子的坐标和动量不能同时确定。牛顿力学的轨道概念在微观领域失去了意义。

**问题之六：微观粒子能量的不连续性。**

按照牛顿理论的观点能量是连续变化的，但在微观领域，实验观测到的原子光谱是分立

的，说明微观领域的粒子能量可以取分立值。牛顿力学及其时空观念，在人类对自然现象的观测和实验中遇到的困难还有很多。

**3. 狭义相对论的时空观**

狭义相对论的两条公理是：狭义相对性原理和光速不变原理。

如果要满足这两条公理，则牛顿的绝对时间概念必须改变。爱因斯坦在1905年的第一篇关于相对论的论文中阐述了相对论的时间概念，其要点是：

其一，以不同速度匀速运动的惯性系中的标准钟的快慢是不同的；

其二，在同一地点对准过的标准钟，可以不同步，与标准钟所在的惯性系和地点有关，也就是同时性不是绝对的；

其三，时间是空间相关的，如果考虑空间中存在引力，时间快慢不但可以不均匀，而且可以有始有终；

其四，这种相对论时钟的快慢变化与具体时钟的计时误差无关，是物理时空本身造成的。

（1）狭义相对论的两条基本原理（假定）

爱因斯坦的狭义相对论是建立在两个基本假设基础之上的。一是相对性原理，即物体运动状态的改变与参考系的选择无关；二是光速不变原理，即对任何一个参考系而言，光速都是相同的。在这两个基本假设中，爱因斯坦抛弃了经典物理学中的“以太”和绝对的时空观。

从这两个基本假设出发，爱因斯坦又推出以下主要结论：①长度收缩，即运动物体在运动方向上长度缩短；②时钟变慢，即运动着的时钟要变慢；③光速极限，即任何物体的运动速度都不可能超过光速；④同时性是相对的，即在一个惯性系中同时发生的事情，在另一个惯性系中测量便不是同时发生的；⑤如果物质运动速度比光速小得多，相对论力学就变为牛顿力学，因此相对论比牛顿力学具有更普遍的意义。

（2）狭义相对论的时空表述

空间和时间的相对性和统一性，是狭义相对论的核心思想。狭义相对论则从数学关系上精确地揭示了空间和时间在本质上的统一性，所谓孤立的空间和孤立的时间在自然界是根本不存在的。后来，爱因斯坦这一新的时空观由他的大学数学老师闵可夫斯基(Minkowski，1864—1909)作了重大发展。他在通常的三个空间坐标以外，又引进了新的第四个坐标，用 $Z=\mathrm{i}ct$ 表示($X$，$Y$，$Z$ 则用 $Z_1$，$Z_2$ 和 $Z_3$ 表示，i是虚数，$c$ 为光速)。从此人们才弄清了，原来自然界的每一真实事件，都只能在四维时空连续区才能做出全面的描述。

**4. 广义相对论的时空观**

爱因斯坦对于1905提出的相对论并不满意，因为其中只涉及了相对作匀速运动的参考系，而没有考虑到加速运动，所以并不完备。这也是今天我们称它为“狭义”相对论的原因。1915年，爱因斯坦进而把相对性原理从匀速运动推广到加速运动，完成了广义相对论的完整表述。1916年写成总结性论文《广义相对论的基础》，宣告了广义相对论的诞生。

（1）广义相对论的两条基本原理（假定）

广义相对论基础的两个基本原理：一是广义相对性原理；二是等效原理。

第一条“广义相对性原理”的物理意义是：所有参考系都是平权的，物理规律具有适用于任何参考系的性质。该原理被称为广义相对性原理。广义相对性原理从根本上否定了特殊参考系，从而否定了牛顿绝对空间。爱因斯坦还利用广义协变的方法，使所有的物理规律

在各个局域惯性系之间变换时具有相同的形式。

第二条“等效原理”的物理意义是：惯性质量×惯性加速度＝引力质量×引力场强度。该原理被称为广义相对论的等效原理。等效原理是广义相对论最重要的基本原理。这个原理的实验依据是由匈牙利物理学家厄缶(Loránd Eötvös，1848—1919)所做的著名的厄缶实验精确证明的引力质量和惯性质量的等价性。

(2) 广义相对论的时空表述

广义相对论实际上是关于空间、时间与万有引力关系的理论，它指出空间—时间不可能脱离物质而独立存在，空间、时间随物质分布和运动速度的变化而变化，揭示了时空同物质的内在联系。广义相对论进一步指出，由于物质的存在，时间和空间会发生弯曲，万有引力实际上是时空弯曲的表现。

如果说现代物理开始于量子物理，经典物理则终结于爱因斯坦的广义相对论。广义相对论的时空观无疑彻底改革了牛顿的时空观。爱因斯坦用相对论的因果律代替了牛顿的绝对时空中的因果律，所以说爱因斯坦的时空概念与因果概念仍然是经典的，广义相对论是经典物理的极致。

在引力场中，在时空的任何一点都可以找到一个局部惯性系(强度不变的范围内)，物理定律在这个局部惯性系中与没有引力场时完全相同。下面从爱因斯坦的引力场方程出发来探讨广义相对论的时空观。

(3) 广义相对论的时空是弯曲的时空

在狭义相对论中，每一个惯性系都对应着闵可夫斯基四维时空中的一个笛卡儿坐标系；不同的惯性系之间的变换，相当于坐标系的一个旋转。如图1-5-1所示，在曲面 $S$ 的 $A$、$B$ 两点各有一个切平面，切平面的法线方向分别沿 $Z$、$Z'$ 轴，并且 $Z'$ 相对 $Z$ 有一定的旋转，相当于闵可夫斯基四维时空中的两个不同的笛卡儿坐标系，即相当于在变化的引力场中的两个时空点 $A$、$B$ 两点的邻域的两个局部惯性系。这样，研究变化的引力场的问题就转化为研究某个曲面 $S$ 的切平面如何连续变化的问题了。换句话说，就转化为研究某个曲面 $S$ 在某个点的切平面和该点的函数关系已知的情况下，求曲面在空间分布的几何问题了。由于时空各点处的引力场强度唯一地确定了时空各点处的弯曲程度，也就可以确定时空各点处的引力场强度。广义相对论和弯曲时空的几何学——黎曼几何形成了对应。

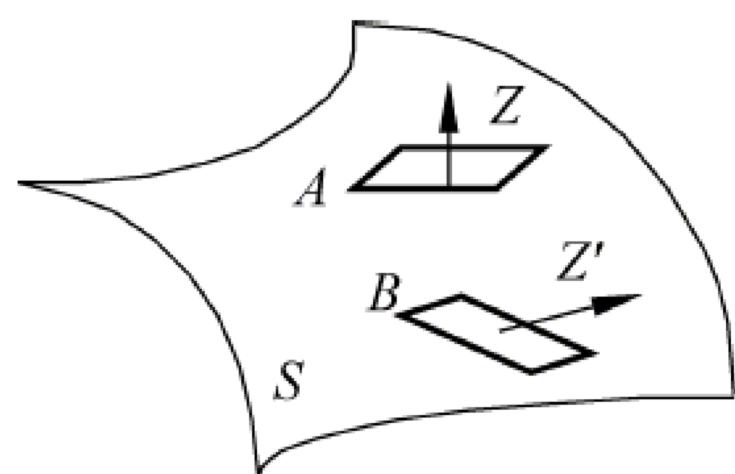


图1-5-1 两个笛卡儿坐标系

(4) 广义相对论的实验验证

实验验证之一：水星的近日点进动。1845年，法国的天文学家勒维列(Urbain Jean Joseph Le Verrier，1811—1877)发现水星的近日点不断前移，广义相对论之前的理论无法给出解释，而广义相对论给出了这一解释。

实验验证之二：广义相对论给出光线在太阳附近偏折的正确数据，与观测结果一致。

实验验证之三：引力红移。光线从大质量的物体射出光谱将向红端移动，反之，光线射向大质量的物体光谱将向蓝端移动。这些结论也与观测结果一致。

总之，广义相对论进一步发展了狭义相对论的时空理论，明确给出了物质及其运动和时空的关系，它描述了在物质平均密度大(引力场强)，宏观物体高速运动的时空特征。关于时空的理论目前还在不断的发展之中，宇宙学的许多谜团还没有解开。

### 1.5.4 爱因斯坦的物理哲学思想

首先,爱因斯坦晚年探索统一场论始终是以时空问题的探索为主线的。在认识时空问题上,他那敏锐的判断力和直觉与早年一脉相承,仍然无人可以企及,统一场论未获结果只能属于时代的局限:即数学表征和实验技术手段的制约;在纪念相对论百年的今天,统一场研究依旧未能取得任何可确认的进展。

其次,爱因斯坦晚年对量子力学是采取排斥态度的。问题应当源于爱因斯坦基于哲学思维上的科学信念对立,即他坚信"上帝不会掷骰子",并非意味着晚年的爱因斯坦削弱了对于科学事物的判断能力。科学界逐渐给世人提供了一个重新认识和发现爱因斯坦的机会。我们只有把爱因斯坦晚年时空观点从其尚未成功的统一场论(亦即纯引力场论)中提取出来加以研究,才能保证完整、准确地认识和理解爱因斯坦和他的相对论。譬如,爱因斯坦生前即已提出时空既是弯曲的也是运动的,可是这么多年过去了,仅就时空弯曲这一个观点而言就远未能被公众所理解和接受,更不要说其他诸多观点了。这种现实客观地表明,爱因斯坦的科学观念不但远远超越他所处的时代,而且也远远超越他身后的未来。

总之,如何对待爱因斯坦晚年关于时空问题的探索与研究,克服某些偏见是相当重要的。人们毕竟已习惯于认为爱因斯坦晚年没有留下有研究价值的理论成果,近年来逐渐重新认识爱因斯坦晚年的时空观点,而且正在取得积极的成果。前不久,关于引力波实验证据的发现就是人类重新认识爱因斯坦的最好例证和回报。

## 1.6 运动与静止

从哲学层面上,辩证法肯定了矛盾为物质本体所固有,并由此构成物质的运动、变化和发展的机制和原因。物质运动的每一既成形态或具体事物都处于生成和灭亡的不断变化中,就物质本体及其"一般的变化"意义上的运动来说,是无始无终、无边无际的,所以说是无条件的、永恒的和绝对的,这就是所谓的"运动的绝对性";而从物质运动的特殊形式或静止来看,则是有始有终、有边有限的,所以说是有条件、暂时的和相对的,这就是所谓的"运动的相对性"。"运动的绝对性"与"运动的相对性"的辩证关系是:变是绝对的,不变是相对的;变和不变、绝对和相对是统一的。

从人类思维的角度理解运动与静止的含义是:运动表现在静止之中,静非不动,而是运动的现实表现或形式;绝对寓于相对之中,相对并不排斥绝对,而是绝对的必然体现或环节。

从物理学层面上,运动与静止的含义是:在承认无条件的、永恒的和绝对的物体运动的同时,物体运动的描述却是相对的,是一种数学方法,即参考系的选择,即物体运动的一种时空描述方法。物体是处于相对的运动还是相对的静止,取决于参考系的选择。

### 1.6.1 运动的含义

所谓运动,即"物质的固有属性"和"存在的方式","它包含宇宙中发生的一切变化和过程"。

**1. 运动的绝对性**

物质是客观存在的,运动不能脱离物质,也不能脱离时空而存在。说到运动,就必须要问,是什么在运动?运动必须有运动的载体,没有载体的运动是不可能存在的。客观唯心主义者(如黑格尔)认为物质之前是"绝对精神"在运动。主观唯心主义者(如马赫)认为运动只

是意识的运动、概念的运动。这种脱离物质的运动是根本没有的。精神的运动、概念的运动是依赖于物质运动的，它是人脑这种高级物质运动的表现形式，是人脑对客观存在的反映。既然运动离不开物质，运动也就不能脱离物质存在的基本形式——时间和空间。离开物质的运动是不可能的，离开时空的运动也是没有的。

如中国古代《墨经》中说："动或徙也。"这是说，运动是空间、地点上的物质发生位移。古希腊的哲学家亚里士多德说："每一运动的事物必有一处所，'非是'原无所处；但它若有运动这便当有其处所。"又说："变于处所是运动。""永恒事物，则凡不生不灭而于空间能运动者亦当具有物质，但这只是空间由一处动向另一处的'运动物质'。"这都是说运动就是物质在时空中的位移。

爱因斯坦表述的定量方程：质能关系 $E=mc^2$ 中，"能"即意味着运动，再次站在物理学的角度从根本上说明了"物质与运动是密不可分"的。

总之，物质存在的绝对性，物质和运动的不可分性，决定了运动的绝对性、无条件性。

**2. 运动的原因**

物质运动的原因是什么？既然运动是物质的固有属性，那么产生运动的原因也必然和物质有关，就是物质的矛盾性。这种原因既包括物质自身内部的矛盾性(这是主要的)，同时也包括一物质与另一物质或多种物质之间的外部矛盾性(这是次要的，但也是必要的)。物质内部与物质外部之间的矛盾性就是造成物质运动的原因。但是造成物质运动的原因到底是一种怎样的矛盾性？我们认真地观察矛盾着的两个方面(差别)发生关系(联系、同一)时，就会发现矛盾着的方面(实体)必然会出现两种作用。这两种作用就是分离性(排斥)与结合性(吸引)。矛盾的分离性与矛盾的结合性是在矛盾的差别性与矛盾的同一性(联系)的实体基础上产生的两种作用，而这两种作用的相互作用正是造成物质运动的具体原因和直接动力。恩格斯说："一切运动都存在于吸引和排斥的相互作用中。"从物理学涉及的运动中举例，拿鸟类的飞行运动来说，鸽子、麻雀从地面上飞起或在空中飞翔时，总是扇动两只翅膀，作上下翻飞的拍打，而这上下翻飞、拍打起伏的动作即是两个翅膀在伸张与收缩、分离与结合，也即在作排斥与吸引的相互作用，更不用说电磁学中关于电磁场中电荷的移动、热学中分子的热运动等。由此可见，矛盾对立双方之间的排斥与吸引的相互作用就是一切客观具体事物产生运动的直接原因。

**3. 运动和静止的关系**

在理解运动时，往往会碰到"静止""绝对""相对"这些概念。所谓静止，是对物质运动的特殊形式或一时状态的表称。所谓绝对，即无条件、永恒性、不可变更之意。所谓相对，则包括有条件性、暂时性、易逝性等内涵。

运动与静止皆统一于物质存在的形式中。没有运动，便没有静止，没有静止也无运动可言。但二者并不是一回事。静止的存在是相对于运动而言的，是运动的特殊形式。静止与运动互相依赖、包含，可以互相易位转化。因此，运动应当从它的反面即从静止之中找到它的量度，获得自身的规定性。因为从辩证的观点来看，运动表现于它的反面即静止之中。运动和静止的关系正如我们已经所述的，静止只是有限运动的结果，而任何有限的运动，都不过是无限的物质运动的一个环节与表现。

### 1.6.2 物理学的运动

通常，人们将物质运动分为五种形式：机械运动、物理运动(包括热运动、电磁运动、原子和原子核内部的运动等)、化学运动、生物运动和社会运动。

物理学只研究低级的运动形式：机械运动和物理运动。它们分别是力学、热学、电磁学、光学、原子核物理学、粒子物理学的研究对象。

**1. 物理学的运动和物理学的变化**

不能把物理学运动与物理学变化混为一谈。运动与变化是分不开的，正如运动与物质是分不开的一样。变化包含运动，但是，变化却不是运动本身所能代替得了的。物理学运动，从准确严格的意义上讲，只指位移，只是物质在总体上所作的单纯位置迁移，但不涉及事物自身在数量上的增减和性质上的转换与变更。但物理学变化却复杂、高级得多。因为变化不单是运动，不仅仅是波动性与旋转性的位移，而且还有自己的独特内容，即它必须在波动性、旋转性位移(运动)时，引起物质自身在数量上的增减与质量上的变更、转换、交易。显然不能简单地一概把这些变化看成是一般性的运动。物理学运动说明不了物质本身在性质上的变换和不同，也不能解释物理世界千差万别的多样性(量)与复杂性(质)。物质自身或物质之间的物理学变化使物质世界千姿百态。物理运动说明不了事物自身性质上的不同与物理世界的复杂性、多样性。只有变化才提供了认识物质本身中的差别与世界的复杂性、多样性的钥匙。

**2. 机械运动**

机械运动指物体占据一定的空间位置，并且随着时间的推移而发生空间位置的改变。观察者也总是参照本身所处的空间位置去描写物质的运动，即描写它的位置及变动。因此要准确地描述运动，需要三个先决条件：选择参考系、确定时间的量度和确定空间的量度。在经典力学范围内，人们认为时空是连续的，这样有利于比较方便地建立时空的实际量度标准，从而对机械运动进行定量的、系统的描述和分析。

(1) 参考系

世界上一切物质都在运动，但一定物质的运动情况如何，则具有相对性。例如火车启动后观察车厢里的物体，车上的乘客和站台上的人的说法就不同，因为乘客选择车厢为标准，站台人员选择地面为标准，一个物体相对于不同的标准具有不同的运动状况。所以我们谈到某个物体是静止的时候总是相对于另一个物体而言：房屋、桥梁相对于地球而静止，乘客相对于车厢而静止等。也就是说，静止是相对的。由于机械运动具有相对性，为了确切说明一个物体的位置和运动，就必须选择其他物体作为标准，描述运动时选作标准的物体，称为参考系。为了定量地描述物体相对于一定参考系的运动，还需要在参考系上固定一种坐标系，例如，加入空间坐标系，可用坐标的改变量描述位置的变动。

(2) 时刻和时间

物质的机械运动是物质的位置随时间的流逝而发生的变动。在这种运动中，物质的每个位置均与一定的时刻相对应。时刻指时间流逝的一刹，又叫瞬时。用数字说明时刻，一要规定时间单位，考虑到时间的均匀性，我们等间隔地划分时间；二要选择某时刻作为计时起点。所谓时间间隔指自某一初始时刻至终止时刻所经历的时间，物体位置的变动总是与一定的时间间隔相对应。

(3) 测量长度和时间的标准

在一定的参考系和坐标系中观察和描述运动，需要用尺对坐标轴进行刻度，需要用时钟测量时刻和时间。这些测量必须有一整套法定的单位标准，具体内容请参阅本书3.2节和3.4节。

**3. 热运动**

分子热运动的基本特征是永恒的运动与频繁的相互碰撞。它与机械运动有本质的区

别,故不能简单应用力学定律来解决分子热运动问题。

(1) 无序性:某个分子的运动,是杂乱无章的,无序的;各个分子之间的运动也不相同,即无序性;这正是热运动与机械运动的本质区别。

(2) 统计性:从大量分子的整体的角度看,存在一定的统计规律,即统计性。

例如:在平衡态下,气体分子的空间分布(密度)是均匀的。(分子运动是永恒的)

可作假设:气体分子向各个方向运动的机会是均等的,或者说,沿各个方向运动的平均分子数应相等且分子速度在各个方向的分量的统计平均值也相等。对大量分子体系的热平衡态,它是成立的。

宏观量:表征大量分子整体特征的量,如温度、压强、热容等,是实验中能测得的量。

微观量:表征大量分子整体中个别分子特征的物理量,如某个分子的质量、速度、能量等,在现代实验条件下是不能直接测量的。

(3) 统计方法:分子热运动具有无序性与统计性,与机械运动有本质的区别,故不能简单应用力学定律来解决分子热运动问题,必须兼顾两种特征,应用统计方法。

统计方法同时伴随着起伏现象。如对气体中某体积内的质量密度的多次测量,各次测量对平均值都有微小的偏差。当气体分子数很大时,起伏极其微小,完全可忽略;当气体分子数较小时,起伏将与平均值可比拟,不可忽略。所以,统计规律只适用于大量分子的整体。

气体动理论中,运用统计方法求出大量分子的某些微观量的统计平均值,用它来解释实验中测的宏观量,故可从实测的宏观量了解个别分子的真实性质。

**4. 电磁运动**

人类有关电磁现象的认识可追溯到公元前600年,在那时古希腊已有关于天然磁铁矿石吸引铁的现象的记录。我国古代更是利用电磁现象做出了指南针等伟大发明。但是,人类对于产生电磁现象的本源即电磁运动的了解却要到18世纪。英国物理学家格雷(Grey,生卒年不详)最先将静电学作为一门独立的学科加以研究。1729年,他在研究琥珀的电效应是否可传递给其他物体时发现导体和绝缘体的区别:金属可导电,丝绸不导电,并且第一次使人体带电。

1745年荷兰物理学家马森布洛克(P. V. Musschenbrock,1692—1761)发明了莱顿瓶。马森布洛克是荷兰莱顿人,莱顿瓶因此而得名。1747年,美国的富兰克林(Benjamin Franklin,1706—1790)根据自己多年的实验得出:电可以像流体一样在物质间流动。人类第一次对电磁运动有了比较深入的了解。从此,对电磁运动的研究更深入、更广泛。1820年奥斯特(Hans Christian Oersted,1777—1851)发现电流磁效应,人类首次证明了电和磁是相互联系的,由此启发人们探索磁生电的可能性。1831年法拉第(Michael Faraday,1791—1867)发现了电磁感应定律,实现人类磁生电的梦想,真正使人类步入了现代文明,同时,法拉第的电磁感应定律也说明:变化着的磁场可以激发涡旋电场,否则,就不会有感应电动势或感应电流。

麦克斯韦将安培环路定理推广到非稳恒电路的情形,说明传导电流是激发涡旋磁场的源泉,提出了位移电流的假说,并进一步用偏微分方程组(麦克斯韦方程组)说明位移电流也是激发磁场的源泉,也就是说,变化着的电场激发涡旋磁场,而这正是产生电磁波的必要条件之一。赫兹的实验证明了电磁波的存在,这就为位移电流的假说提供了最有力的依据,也充分说明了电磁运动的存在和重要地位。

**5. 原子和原子核内部的运动**

古代哲学家早对物质的结构有过许多设想,认为物质是由简单的、不可分割的基本单元

所谓“原子”构成，这是原始的原子学说，建立在科学基础上的原子学说是到了近二三百年内才发展的。到了19世纪末，原子物理学才得以快速发展，到了1925年一个关于微观体系的新理论——量子力学——建立了起来。

在整个原子物理学的发展过程中，1827年英国著名植物学家罗伯特·布朗(Robert Brown，1773—1858)完成了对分子运动的证明；1833年法拉第电解定律的提出显示了电的基本单元(电荷)的存在；1885年巴耳末(Johann Jakob Balmer，1825—1898)发现氢光谱线系规律；1887年赫兹(Heinrich Rudolf Hertz，1857—1894)发现光电效应；1895年伦琴(1845—1923)发现X射线；1896年贝克勒耳(Antoine Henri Becquerel，1852—1908)发现放射线；1897年汤姆孙证明电子的存在；1900年普朗克(Max Karl Ernst Ludwig Planck，1858—1947)进一步研究黑体辐射的结果以后，提出了量子论；1911年卢瑟福证实了原子的核式结构，在此基础上1913年玻尔理论发表。量子力学的建立，圆满地解决了氢原子的结构及其光谱问题，泡利不相容原理圆满地解释了元素周期律原子能级的壳层结构。这一连串的科学发现是在研究原子及其内部物质运动的过程中积累的，一步步地揭示了原子和原子核的内部结构，反过来对于微观物质结构的了解又有力地促进了对于微观运动的研究。

## 学习资源

### 一、中国近现代著名科学家(物理学)简介

钱学森(1911—2009)，出生于上海，浙江杭州人，中国空气动力学家，中国科学院、中国工程院院士，中国两弹一星功勋奖章获得者之一。“工程控制论”创始人，为中美两国的导弹和航天计划都曾做出过重大贡献。



钱学森，英文名：Qian Xuesen、Tsien Hsue-shen、H. S. Tsien。别名：Qian Xuesen。毕业于上海交通大学机械工程系，后作为清华大学留美公费生到美国麻省理工学院、美国加利福尼亚理工学院学习。被誉为“中国现代火箭之父”的钱学森，是著名的航天工程和空气动力专家。他早年留学美国，在冯·卡门教授的指导下，在火箭研究中取得了重大进展，为反法西斯的胜利做出了重大贡献。1947年，刚刚36岁的钱学森被聘请为美国麻省工学院的终身教授。新中国成立的喜讯传到钱学森那里，他想：“我是一个中国人，我可以放弃这里的一切，但不能放弃祖国。我应早日回到祖国去，为建设新中国贡献自己的全部力量。”为了报效新生而落后的祖国，钱学森从1950年起在美国向其当局正式提出回国申请。但是，美国当局却百般阻挠并加以迫害，没收了钱学森的各种资料和书籍，并诬蔑陷害他为“间谍”，对他进行审讯和监禁，将他关押在一个孤岛上，仅半个月，就使他的体重减轻了14kg。当时美国当局声称，只要钱学森放弃回国念头，就照常给他提供实验室和仪器设备。可是，钱学森宁死也要回国，始终没有屈服。通过五年的艰苦斗争，在周恩来总理的亲切关怀下，钱学森于1955年9月17日踏上了归国的路程。回国后，钱学森为新中国的航天事业跃入世界前列立下了不朽的功勋。

## 二、诺贝尔物理学奖简介

诺贝尔1833年生于瑞典的斯德哥尔摩,诺贝尔一生致力于炸药的研究,在硝化甘油的研究方面取得了重大成就。他不仅从事理论研究,而且进行工业实践。他一生共获得技术发明专利355项,并在欧美等五大洲20个国家开设了约100家公司和工厂,积累了巨额财富。

**1. 奖项设立**

1896年12月10日,诺贝尔在意大利逝世。逝世的前一年,他留下了遗嘱,设立诺贝尔奖。据此,1900年6月瑞典政府批准设置了诺贝尔基金会,并于次年诺贝尔逝世5周年纪念日,即1901年12月10日首次颁发诺贝尔奖。自此以后,除因战时中断外,每年的这一天分别在瑞典首都斯德哥尔摩和挪威首都奥斯陆举行隆重的授奖仪式。诺贝尔奖最初分设为物理、化学、生理学或医学、文学、和平五个奖项。

1968年瑞典中央银行于建行300周年之际,提供资金增设诺贝尔经济奖(全称为瑞典中央银行纪念阿尔弗雷德·伯恩德·诺贝尔经济科学奖金,亦称纪念诺贝尔经济学奖),并于1969年开始与其他5项奖同时颁发。诺贝尔经济学奖的评选原则是授予在经济科学研究领域做出重大贡献的人,并优先奖励那些早期做出重大贡献者。

1990年诺贝尔的一位重侄孙克劳斯·诺贝尔又提出增设诺贝尔地球奖,授予杰出的环境成就获得者。该奖于1991年6月5日世界环境日之际首次颁发。

**2. 颁奖时间**

每次诺贝尔奖的发奖仪式都是下午举行,这是因为诺贝尔是1896年12月10日下午4:30去世的。为了纪念这位对人类进步和文明做出过重大贡献的科学家,在1901年第一次颁奖时,人们便选择在诺贝尔逝世的时刻举行仪式。这一有特殊意义的做法一直沿袭至今。

**3. 奖金奖章**

诺贝尔奖的奖金数视基金会的收入而定,其范围约从11000英镑(31000美元)到30000英镑(72000美元)。奖金的面值,由于通货膨胀,逐年有所提高,最初约为3万多美元,19世纪60年代为7.5万美元,80年代达22万多美元。

金质奖章重约半磅,内含黄金270g,奖章直径约为6.5cm,正面是诺贝尔的浮雕像。不同奖项,奖章的背面饰物不同,每份获奖证书的设计也各具风采。颁奖仪式隆重而简朴,每年出席的人数限于1500~1800人;男士穿燕尾服或民族服装,女士要穿严肃的晚礼服;仪式中所用白花和黄花必须从圣莫雷空运来,这意味着对诺贝尔的纪念和尊重(因为圣莫雷是诺贝尔逝世的地方)。

**4. 评选过程**

每年9月至次年1月31日,接受各项诺贝尔奖推荐的候选人。通常每年推荐的候选人有1000~2000人。

具有推荐候选人资格的有:诺贝尔奖获得者、诺贝尔奖评委会委员、特别指定的大学教授、诺贝尔奖评委会特邀教授、作家协会主席(文学奖)、国际性会议和组织(和平奖)。

不得自荐。

瑞典政府和挪威政府无权干涉诺贝尔奖的评选工作,不能表示支持或反对被推荐的候选人。

每年2月1日起,各项诺贝尔奖评委会对推荐的候选人进行筛选、审定,工作情况严加

保密。

每年 10 月中旬，公布各项诺贝尔奖获得者名单。

每年 12 月 10 日是诺贝尔逝世纪念日，在斯德哥尔摩和奥斯陆分别隆重举行诺贝尔奖颁发仪式，瑞典国王出席并授奖。

**5. 推荐流程**

根据规定，下列人员有权推荐诺贝尔物理学奖获奖人选：

(1) 瑞典皇家自然科学院院士或外国院士；

(2) 诺贝尔物理委员会的委员；

(3) 曾被授予诺贝尔物理学奖金的科学家；

(4) 在乌普萨拉、隆德、奥斯陆、哥本哈根、赫尔辛基大学、卡罗琳医学院和皇家技术学院永久或临时任职的物理教授，以及在斯德哥尔摩大学有永久性职务的物理学教员；

(5) 出于使各国和它们的学术中心能够得到相宜名额分配的考虑，由皇家自然科学院选择至少六所大学或具有同等水平的学院中的教授，担任同类职务的人员；

(6) 自然科学院认为可能合乎邀请目的的其他科学家。

## 思考与练习

1. 论述物理学与大学物理学的研究范围。
2. 物理学的概念范畴是如何划分的？
3. 实物与场的区别是什么？
4. 粒子与波是不同的物质吗？试从经典物理学的观点和现代物理学的观点作简单论述。
5. 介质波与概率波的本质区别是什么？
6. 如何理解时间的单向性？
7. 物理学时空观的发展演变给我们什么启示？
8. 试从科学思维的角度论述“运动与静止”的辩证关系。

第2章

# 物理学方法论

我们知道,物理学是一门以实验为基础的定量科学,所以,在观察与实验方法的基础上,运用数学方法进行定量演绎探讨,运用逻辑方法进行思维提炼和升华,从而建立物理学的理论体系。而进行的科学理论思维则贯穿于物理学研究的始终,科学思维的具体方法则随着物理学的发展而日趋丰富。系统地了解物理学研究方法,具体地认识物理学理论是怎样概括总结出来的,对物理学的教学和研究工作有着启发和帮助的作用。

本章系统地介绍了物理学方法论的基本概念、研究的对象和基本内容,讨论了物理学方法论的作用和意义。

## 2.1　物理学方法论概述

虽然物理学的研究对象是物质最低级的运动形式,但是其研究途径却是一种高度复杂的实践与思维过程。随着研究的深入和物理学的广泛应用,物理学已经形成许多新的分支,由此随之产生了许多特殊的研究方法。

### 2.1.1　方法与方法论

自从人类进入文明社会以来,对方法这一概念有不完全相同的定义和理解。“方法”一词起源于希腊词“μετα”(沿着、顺着)和“οδοσ”(道路),它的字面意义是沿着(正确的)道路运动。

**1. 方法**

所谓方法,就是为了解决某一具体问题从实践或理论上所采取的手段的总和。

方法起源于人类的实践活动。人类通过方法这种工具与客观发生关系,所以,方法是属于主观范畴的。例如,日月运行,昼夜交替,这些存在的本身是无方法可言的,但是,我们要认识它们就要涉及方法。而且,不同的人去解决同一问题往往会有不同的方法。例如,测一圆周的长,可以用绳子沿圆周绕一圈然后测绳长;也可以用小滚轮沿圆周滚一周,用滚动的圈数乘以小滚轮的周长;还可以测出它的直径用计算的方法求出周长等。

众所周知,美国试爆第一颗原子弹时,费米想亲自测定原子弹爆炸的威力。于是他将一把事先准备好的纸片抛向空中,然后根据自己离开爆炸中心的距离和纸片被冲击波吹过的距离,迅速推算出原子弹爆炸的威力,计算结果竟然和仪器测量结果相差无几。当然,要是他缺乏有关专业知识,就难以进行这样的计算。这也说明,目标相同,方法可以不同。只要潜心研究,就能找到简单而合理的新方法。

同一事物重组变序以获得不同的结果,也可称为方法。例如,战国时田忌和齐威王赛马,分上、中、下三等一一对应比赛,由于田忌的马的能力不如齐威王的,因而连负三局。此时孙膑向田忌献策:以下等对其上等,宁负一局,然后以上对其中,以中对其下,连胜两局,

以二比一获胜。此法可谓妙哉！

被实践检验过的科学理论知识，当用来在其知识领域内或其他知识领域内建立其他理论时，就实质来说，也起方法的作用。而且往往抽象程度较高的知识对较为具体的知识发挥着方法的功能。所以，从这个意义上来讲，一切知识都可以应用而转化为方法。例如，控制论在研究电子计算技术时就起着方法的作用；极限是数学中的基础理论知识，当用它来建立瞬时速度或瞬时加速度等概念时，就成为极限方法了。而方法一旦在往日的研究结果中形成，就会成为日后研究的出发点。

**2. 方法存在的形式**

（1）对于同一事物来说，沿纵向或横向发展过程中的转折过渡处必然存在方法。

例如，从部分电路欧姆定律出发，研究全电路欧姆定律时，必然要用到实验归纳与理论演绎相结合的方法。

（2）不同事物之间（包括人和事物之间）建立联系或者发生关系时，必然存在方法。

例如，使闭合回路的一部分导体在磁场中作切割磁力线运动，或者使闭合回路的磁通量发生变化，运用上述方法都可以使磁与电这两种事物建立起联系。

（3）理论用于实践解决问题时，理论本身就具有了方法的意义。

例如，研究一个物体从光滑斜面顶端下滑到底部时所具有的速度，我们可以用机械能守恒定律，或者牛顿第二定律与运动学公式相结合来求解。此时，上述理论实际上就成为解题方法了。

还应强调的是，当新的科学理论建立时，往往会引起思维的变革。例如量子力学的建立，导致以统计因果观为核心的思维方式取代了以严格决定论为核心的经典思维方式。

**3. 方法论**

1）什么是方法论

方法论是与希腊词 λονοσ 有关的一个概念，它是关于认识和改造现实方法的学说和理论。就词的本义而言，方法论一词的创始人是英国哲学家培根。他首先提出以方法论体系武装科学思想，并在《新工具》一书中付诸实现。他对科学认识的归纳法及应用法所做的论证，对后来方法论的发展都起了重大的作用，同时使方法论问题成为哲学的中心问题之一。

方法论知识既可以是某些规定和标准的形式，用以确定某些特定活动类型和顺序（标准方法论），也可以是实际已经完成的某一活动的描述形式（描述方法论）。

2）方法论的分类

由于标准不同，方法论可以有许多不同的分类方式。按照方法论分析的不同层次或普遍性程度的高低来划分，它包括既相互联系又相互区别的三个层次。

（1）哲学方法论。也就是辩证法、认识论和辩证逻辑，它普遍适用于自然科学、社会科学和思维科学。比如一切从实际出发的方法、矛盾分析方法等。

（2）自然科学方法论。它是从自然科学的各门学科中概括出来的，诸如观察和实验的方法、抽象思维和形象思维的方法、数学方法等。

（3）各门学科中的一些具体方法。如初等数学中的数学归纳法，高等数学中的微分法、积分法，物理学中的光谱分析法，化学中的催化方法等。

### 2.1.2　物理学方法论的研究对象和内容

物理学方法论是以唯物辩证法为指导探讨物理科学一般研究方法的理论。以上各类方法在物理学方法论中几乎都会涉及。它主要是探讨用什么方法研究物理现象，怎样描述物

理现象，怎样探索并总结物理规律等。它既具有自然科学方法论的一般特征，又反映物理科学研究方法的特殊规律。一方面它把哲学方法、自然科学的一般方法与物理学科的特点相结合，运用于物理运动规律性的研究中，另一方面又把物理学科中某些共有的、通用的方法，通过抽象、概括加以提炼，使之上升为自然科学研究的一般方法。

具体说来，物理学方法论包括以下一些内容：

(1) 物理科学认识的逻辑结构和研究程序，揭示物理科学研究过程的各个阶段和每一环节的作用、特点及其所应遵循的一般原则。

首先通过对物理系统的分析，研究一般物理系统，从确定课题到总结规律需要经过哪些环节和步骤，这些活动有什么规律性，要运用哪些方法，这些方法有何作用及特点。其次在物理科学研究的基础上总结出物理学方法论，并从整体上分析其逻辑结构体系。也就是说，分析物理学方法论应由哪些方法组成，这些方法之间的区别与联系又是什么？

(2) 总结物理科学研究中常用的一般方法，并将它们分类，揭示各种方法的含义、特点、适用范围，运用的原则和注意事项，以及物理学史上的例证。并尽可能地给出使用模式，以便使用、借鉴与移植。

例如，物理学方法大致可分为两大类，一类是具有一定程式和规律的常规方法，如观察、实验方法、数学方法、理论思维等方法；另一类是非常规的方法，诸如直觉、灵感、机遇、猜测等方法。

(3) 研究物理学史上的重大突破和有代表性的事例，揭示著名物理学家的研究方法。

物理学方法论的全部内容，包括总结科学方法、探讨科学方法的原理与结构、总结物理学研究的程序等都是从物理学史的反思中以及人们对物理研究过程的再认识中总结出来的。实践证明，重大科学理论的突破，往往伴随着科学方法的诞生。以狭义相对论建立为例，在爱因斯坦以前，洛伦兹与彭加勒(Jules Henri Poincaré，1854—1912)在物理概念及数学形式上都十分接近了狭义相对论，但他们只是限于对牛顿理论的修修补补，极力维护绝对时空观的旧有框架。为什么只有到爱因斯坦时才能最后提出新理论呢？这里面有个方法论的问题。爱因斯坦本人曾指出，英国哲学家休谟(David Hume，1711—1776)与奥地利物理学家马赫(Ernst Mach，1838—1916)的怀疑论对他影响很大，使他敢于对牛顿理论的庞大体系产生怀疑，并树立起推翻牛顿理论的信心。另外，荷兰哲学家斯宾诺莎(Baruch Spinoza，1632—1677)的唯理论方法也给了爱因斯坦建立新理论具体方法的手段，他采用斯宾诺莎的方法建立了公理化的相对论体系，而且爱因斯坦本人还在此基础上提出了具有方法论意义的逻辑简单性原理。

物理学史中还有大量具有方法论意义的思想有待我们去进一步发掘和整理。例如，在量子理论的发展过程中，以玻尔(Niels Henrik David Bohr，1885—1962)为首的哥本哈根学派与爱因斯坦的争论极富方法论意义，特别是玻尔提出的对应原理和互补原理，都有待于从科学方法论的角度去进行总结提高。

应该强调的是，我们不仅要重视物理学家成功的经验，还要注意他们失败的教训，揭示失败的方法论意义，把失误转化为成功，实践证明这也是物理学方法论中不可缺少的内容。

(1) 研究新兴科学、新兴技术对物理学研究的重大影响，并探讨其方法论意义。

20 世纪以来崛起一大批新兴学科、新兴技术，必然对物理学研究方法的发展产生十分深远的影响。例如，20 世纪 40 年代几乎同时出现的控制论、信息论和系统论这三门横向科学，为现代科学技术及物理学的研究提供了崭新的思维方法和从整体上认识事物的系统科

学方法。它们不但是现代科学不可缺少的研究方法，同时也彻底改变了当代科学家的思维方式。总之，新的思维方式与新的技术手段相结合，必然会对物理学发展起到不可估量的作用。

## 2.2　实验方法

众所周知，物理学是一门以实验为基础的科学，一切理论都要以实验作为唯一的检验基准。人们对物理问题总是在观察、实验的基础上，经过一系列的科学抽象，从现象深入到本质，从感性上升到理性，最后形成物理理论；而物理理论在经过可重复科学实验的反复检验后，又进一步系统化而形成一门学科，这就是物理学。追根溯源，可以看到，物理学最开始源于观察与实验，而又必须经受实验的检验，而观察与实验的原动力是人们的生产实践，是人们追求探索自然的奥秘。认识到这一点，才能从本质上理解物理学。实验也是研究和学习物理学最基本的方法。

### 2.2.1　实验方法及其特点

**1. 什么是实验方法**

实验方法是人们根据一定的目的和计划，利用仪器、设备等物质手段，在人为控制、变革或模拟自然现象的条件下获得科学事实的方法。

**2. 实验方法的特点**

(1) 实验方法可以简化、纯化以至强化自然过程

由于任何一个自然事物和生产过程都是非常复杂的，不仅自身表现出各种各样的相互交织的现象，而且还同周围环境相互作用，相互影响。在这种状态下，单靠观察方法难以揭示各现象之间的本质联系。而在实验过程中，科学家可以通过一系列手段干涉自然过程，使之简化、纯化或强化，以便突出主要因素，排除次要、偶然因素，加深对事物实质的认识。从这个意义上看，实验过程本身就是一种抽象，是仪器中所凝聚的思维对自然的抽象。

在物理学中，发现负质子的过程，就是一个很好的简化和纯化过程的例子。当安德森发现正电子以后，物理学家根据对称原理推测，可能存在有负质子(在此之前，一般只有负电子和正质子被人所知)，于是制造了几十亿电子伏特的回旋加速器。但是要观测到负质子，必须能够把同负质子伴随产生的其他粒子滤掉，这样才能找到负质子。于是，美国科学家张伯伦(Owen Chamberlain，1920—2006)和西格雷(E. G. Segre，1955—　)等人借助一种复杂的电磁场、狭缝等构成的“迷宫”，使得当物质遭到轰击时，只有负质子才能穿过“迷宫”达到观察端。这两位科学家通过这种纯化装置，发现了负质子，并于 1959 年获得了诺贝尔物理学奖。在强化过程中，人们可以发现常态下所不能发现的东西。比如，千百年来人们对物质状态的认识只局限于日常所能感知的“固、液、气”三态上。后来，人们在超高温的条件下发现了“等离子态”。主要是因为在超高温的情况下，原子核外的电子能量增大，以至脱离轨道，变成自由电子，于是原子变成了离子。这样物质处于带电离子、电子和未电离的中性原子共有的“等离子态”。同样在高压的情况下，电子被压到原子核中，于是形成“超固态”。

可见，实验中的简化、纯化和强化，可以给我们带来许多新的发现。

(2) 实验可以模拟自然过程

模拟是实验中更为重要和普遍的过程。它可以对不可接近的自然现象进行间接研究，从而扩大研究对象的时空范围。

模拟过程可以减少许多实际实验所带来的麻烦，从而更利于控制自然过程，给研究者带来方便。当然，任何模拟都不是随意的，而是根据相应的相似原理进行。

(3) 在实验过程中要善于把握机遇

科学虽然是人们有目的、有计划的研究活动，但它毕竟是对未知自然现象的探索，因而不可能预先把一切过程囊括无遗，总会不时遇到一些意外现象，而深入研究这些意外的现象又往往导致重大科学发现。因此，在实验过程中要注意捕捉机遇。“机遇只偏爱那些有准备的头脑。”

## 2.2.2 物理实验及其特点

**1. 什么是物理实验**

物理实验是人们根据研究和学习的目的，利用物理仪器和设备，人为地控制或模拟物理现象，排除各种偶然因素的干扰，突出主要因素，在有利的条件下重复地研究物理现象及其规律。物理实验是建立和检验物理理论的基础，是物理学工作者的一种重要研究方法，同时也是学生学习物理的一种基本方法和途径。

下面是用凸透镜做热辐射实验。

(1) 取一张白纸(大小不限)，用墨汁将一半涂黑待干。如图 2-2-1 所示。

(2) 实验：将凸透镜放于太阳光之下，找好焦点(使太阳光在凸透镜下面会聚的亮点最圆最小)，然后固定好凸透镜的位置。先将(1)里的白纸放在凸透镜下的焦点位置上，并启动钟表计时，过很长时间纸才被烧焦。再将涂黑的纸移到凸透镜的焦点位置上，同样启动钟表，过不多时，则见黑纸先冒烟，后起火星，黑纸被烧了个洞。

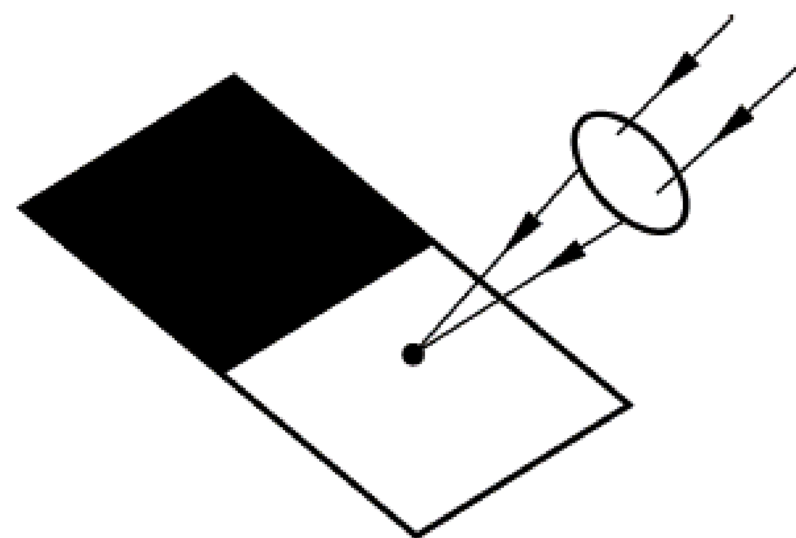


图 2-2-1 用凸透镜做热辐射实验

实验表明：黑色表面的物体比白色表面的物体吸热快。

当然这只是个很简单的小实验，随着研究对象复杂程度的增加物理实验也会更复杂、高级。

**2. 物理实验的特点**

(1) 物理实验能够使研究对象以较为纯粹的状态出现。利用各种手段将研究对象从复杂的自然联系中分离出来，人为地排除各种偶然因素的干扰，使一些现象发生，另一些现象不发生；使一些条件发生变化，另一些条件不变化。其结果使研究对象的运动变化以较为纯粹的状态出现。

(2) 物理实验可以强化观察对象的条件。通过实验手段创造出超高压、超真空、超磁场、超低温等自然状态下难以出现的特殊条件，就能进一步认识新的现象、新的特征。例如，在超低温条件下，可以发现几乎所有气体液化的事实。

(3) 物理实验可以使观察对象重复出现。自然现象受时空限制，有的周期过长，给观察带来困难。利用实验手段可使观察对象在短时间内重复出现，给学生增加观察的机会，以获得更多的感性认识。这是物理实验的突出特点。

**3. 物理实验的目的**

(1) 通过观察、测量和分析，加强对物理概念和理论知识的理解，更加牢固掌握所学过的物理知识。

(2) 学习物理实验的基本知识、基本方法，培养基本的实验技能。要做好一个实验，除

了要了解有关的理论外，还必须能运用恰当的实验方法，合理地选取符合实验要求的仪器，懂得怎样装配、调整及正确操作这些装置，在取得必要的数据后，能从中得出切合实际的结论，并能分析判断实验结果的可靠程度和存在的问题。

(3) 培养严肃认真、实事求是的科学态度和工作作风。

### 2.2.3　物理实验的类型

物理实验种类很多，不同的角度有不同的划分标准。按实验中研究问题的质或量来划分，有定性实验或定量实验；按实验在认识过程中的作用来划分，有析因实验、对照实验、中间实验；从研究的领域来划分，有力学实验、光学实验、电磁实验、基本粒子实验等；有些实验目的在于验证理论上的某些推测，称为验证性实验；有些实验需要模拟自然过程或生产过程中某些物理现象进行研究，可称为模拟实验。下面我们着重介绍一下定性实验、定量实验和验证性实验。

**1. 定性实验**

(1) 什么是定性实验

在实验中对研究的问题不需要或者不可能作量的测定，目的在于弄明白它的物理属性或规律，这类实验都可称为定性实验。在物理学研究中，定性实验是不可缺少的。在物理学发展初期，人们还处于搜集材料、描述和记录材料的阶段，实验手段和工具都非常落后，特别是受知识积累的限制，往往很难对复杂的物理现象作量的研究，所以，定性实验被广泛应用。

(2) 定性实验可以用来判定某些物理现象是否存在及其特性

1921 年 4 月，劳厄(Max Theodor Felix Von Laue，1879—1960)让伦琴射线通过硫酸铜晶体，看到晶体后面的胶片呈现排列规则的感光点，因而判定伦琴射线通过结晶体能够发生衍射，对波长极短的伦琴射线来说，原子的间距就像不透明的“屏”上开了许多窄缝，因此，呈现出衍射图像。爱因斯坦称赞这个实验是“物理学史上最美妙的一页”。劳厄的实验不需要作量的测定，它只是定性地揭示了晶体内原子间的间距与伦琴射线的波长属于同一数量级，是一个较典型的定性实验。

(3) 定性实验还可用以弄清某些物理现象的成因或规律

关于雷鸣电闪现象，古代有人认为，“雷”和“电”都是一种可燃性的雾(繁体字的“电”是“雨”字头)，会慢慢地上升，到一定的时候就以闪电的形式冲破云层表现出来。笛卡儿(Descartes，1596—1650)认为雷电是上层的云下降碰到下层的云，两者碰撞而引起。欧拉(Leonhard Euler，1707—1783)最先认识到雷与“电”有相同的性质，但是一直缺乏可靠的证明。富兰克林在 1752 年 9 月再次利用风筝把云层的电引入室内，进行室内雷鸣闪电实验，证实雷电与电火花放电有同一本质，找到了雷电的成因。

下面附小孔成像的定性实验。

**附：**

用一支削得很尖的铅笔，在一张硬纸片的中心部分扎一个小孔。孔的直径为 3mm 左右。设法把它直立在桌子上(图 2-2-2)。然后拉上窗帘，使室内的光线变暗。

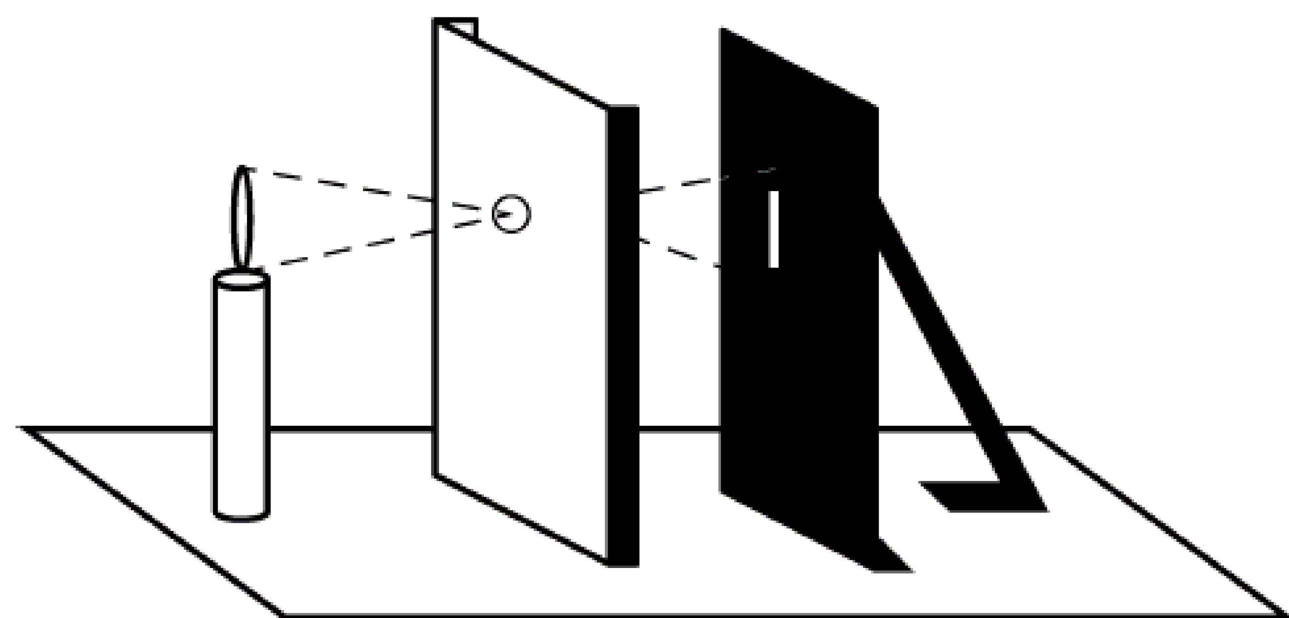


图 2-2-2　小孔成像实验图一

点上一支蜡烛，放在靠近小孔的地方。拿一张白纸，把它放在小孔的另一侧。这样，你就会

在白纸上看到一个倒立的烛焰。我们称它是蜡烛的像。前后移动白纸,瞧瞧烛焰的像有什么变化。当白纸离小孔比较近的时候,像小而明亮;当白纸慢慢远离小孔的时候,像慢慢变大,亮度变暗。

改变小孔的大小,我们再来观察蜡烛的像有哪些变化。

你可以在硬纸片上,扎几个大小不等、形状不同的孔,孔和孔之间相距几厘米。这时候在白纸上,就出现了好几个和小孔相对应的倒像。它们的大小都一样,但是清晰程度不同,孔越大,像越不清楚。孔只要够小,它的形状不论是方的、圆的、扁圆的,对像的清晰程度都没有影响。

这个实验至少向我们提出了三个问题:小孔成的像为什么是倒立的?像的大小和哪些因素有关?像的清晰程度和哪些因素有关?

为了说明这些问题,我们把蜡烛的火焰看成是由许多小发光点组成的,每个发光点都向四面八方发着光。总会有一小束光,笔直地穿过小孔,在白纸上形成一个小光斑。烛焰上的每一个发光点都会在白纸上形成一个对应的光斑,全部光斑在白纸上就组成了一个烛焰的像。

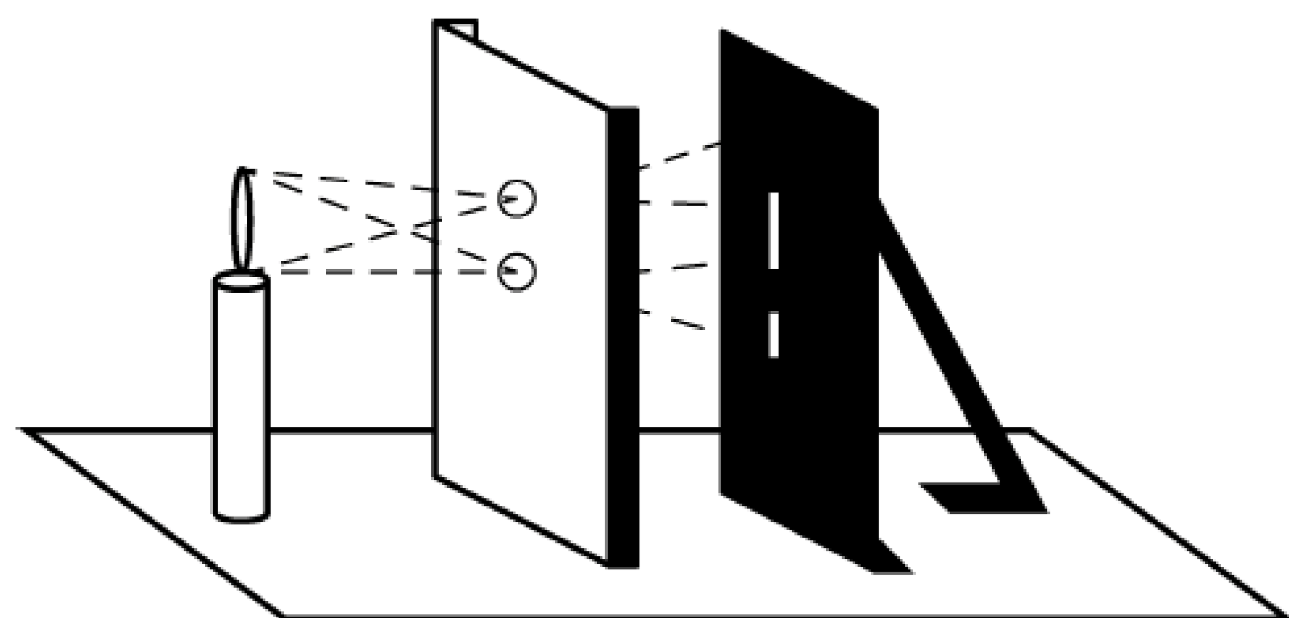


图 2-2-3 小孔成像实验图二

从图 2-2-3 中可以看出,烛焰上部发的光沿直线通过小孔,照在白纸的下部;烛焰下部发出的光,通过小孔,照在白纸的上部,所以在白纸上形成一个倒立的像。这正好说明了光是沿直线传播的。

两千多年以前,我国学者韩非,在他的书里记载了一个有趣的故事:有人请了一个画匠为他画一张画。三年以后,画匠告诉他:“画成了!”他一看,八尺长的木板上只涂了一层漆,什么画也没有,便大发脾气,认为画匠欺骗了他。画匠说:“请你修一座房子,房子要有一堵高大的墙,再在这堵墙对面的墙上开一扇大窗户。把木板放在窗上,太阳一出来,你在对面的墙上就可以看到一幅图画。”他半信半疑,照画匠的话去办。果然,在屋子的墙壁上出现了亭台楼阁和往来车马的图像,好像一幅绚丽多彩的风景画。尤其奇怪的是,画上的人和车还在动,不过都是倒着的!(见图 2-2-4。)



图 2-2-4 画匠说画

大约两千四五百年以前,我国的学者墨翟和他的学生,做了世界上第一个小孔成倒像的实验,解释了小孔成倒像的原因,指出了光的直线行进的性质。这是对光直线传播的第一次科学解释。

**2. 定量实验**

在实验中对研究的问题需要做出精确的数量测定，确定物理现象各种具体的参数、各现象之间具体的数量关系，或者需要用数量去表明某些规律。这类实验都可称为定量实验。它同定性实验一样，在物理学研究中是不可缺少的。随着物理研究的深入，实验手段和工具的发展，人们知识的增多，对定量实验的应用就越来越广泛，定量实验的地位越来越重要。

物理学中长度、质量、密度等物理量的测量；研究电流强度与电阻、电压的定量关系；研究力、质量、加速度的关系等实验都属于定量实验。

**3. 验证性实验**

在物理学研究中，当人们对研究对象有了一定认识之后，根据已知的理论和实验，对一些物理现象的存在、原因或规律做出推测、提出假说，或形成新的理论时，为了检验它们正确与否而设计的实验叫验证性实验。

验证性实验在物理学研究中同样是不可缺少的。在物理学的发展过程中，这类实验的应用都极为广泛，如1955年莱茵斯和科旺证实了中微子存在的实验，就是验证性实验。他们按特定的方式，把中子计数器和电子计数器连接起来，以保证它们同时被中子和正电子打中时才发出信号。因此，只要发出信号，就意味着中子和正电子同时打中按特定方式连接起来的中子计数器和电子计数器，证实了中微子与质子碰撞的过程，也证实另外中微子的存在，他们的实验是成功的。验证性实验可以分为两类：一类是直接验证，如赫兹通过实验直接证实了电磁波的存在和传播；另一类是间接验证，即不去验证理论本身，而是验证其推论，如对广义相对论的验证就是验证其推论的，前文所述的莱茵斯和科旺的实验也属于间接验证实验。

中学物理教材中的学生实验大多是验证性实验，如验证牛顿第二定律，验证动量守恒定律，验证机械能守恒等。

### 2.2.4 物理实验设计的方法

在物理学史上有一些实验设计非常巧妙，不仅目的明确，思考周密，而且能抓住关键，出奇制胜。例如卢瑟福于1911年设计的 α 粒子散射实验，后来根据实验的结果，卢瑟福否定了汤姆孙模型并提出了原子的核式模型。

实验装置如图2-2-5所示。散射源R中发出一细束 α 粒子，直射到金属箔上以后，由于各 α 粒子所受金属箔中原子的作用不同，所以沿着不同的方向散射。闪烁屏S及放大镜M可以沿着以散射箔F为中心的圆弧移动。当S和M对准某一方向上，通过F而在这个方向散射的 α 粒子就射到S上而产生闪光，用放大镜M观察闪光，就能记录下单位时间内在这个方向散射的 α 粒子数。从而可以研究 α 粒子通过散射箔后按不同的散射角 $\theta$ 的分布情况。α 粒子散射实验否定了汤姆孙的原子模型，根据实验结果，卢瑟福于1911年提出了原子的核式模型。原子中心有一个极小的原子核，它集中了全部的正电荷和几乎所有的质量，所有电子都分布在它的周围。卢瑟福根据设想的模型，从理论上推导出散射公式，并被盖革-马斯顿实验所验证，核式模型从而被普遍接受。

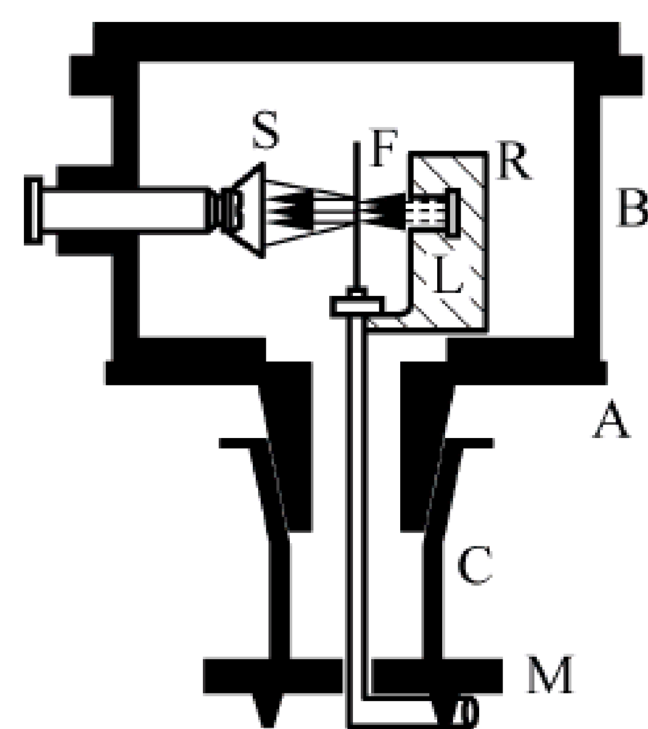



图2-2-5 检验卢瑟福散射公式的实验装置

A—刻度圆盘；B—金属匣；C—光滑轴套；F—散射箔；R—散射源；S—闪烁屏；M—望远镜

从这个例子中可以看出，一个完整的物理实验大致可分

为三个组成部分：

其一，实验源：它是实验信号的发生源，如 α 粒子散射实验中的 α 粒子散射源 R。

其二，实验对象：它是实验源信号所作用的对象，如 α 粒子散射实验中的散射箔 F。

其三，实验效果显示器：它是用以呈现实验对象接受实验源信号作用后所产生的效应部分，以便通过直接或间接的方式进行效果观察，α 粒子散射实验中的闪烁屏 S。

当然，上述分法不是绝对的。有些简单的实验中，实验源和实验对象是一个；有的实验中，实验效果就显示在实验对象身上。

我们了解实验的基本组成部分是为了根据要求进行实验设计。在物理实验中，常用的设计方法有以下几种：

(1) 平衡法：所谓平衡就是矛盾双方的平衡，分析平衡就是分析矛盾的双方。在一个平衡系统中，总存在着偏离平衡的因素，平衡就是偏离平衡的因素的抵消效应，其实质就是用已知的物理量去检测未知的物理量。许多测量仪器就是根据平衡的原理设计的，如天平、比重计、压强计、温度计等。当然这些仪器在使用时也是利用平衡法。一些物理实验如电桥测电阻、测金属的比热容等也都是采用平衡法。

(2) 转换法：也称间接测量法。在物理实验中，常有一些现象因不明显而不易于观察或者不易直接观察，这就要借助于力、热、光、机械等方法之间的相互转换以间接地实现可观察、容易观察或观察效果明显的目的。如麦克尔孙干涉仪把测量光速转换成测光的干涉条纹；伏安法测电阻是把测电阻转换成测电流和电压。

(3) 放大法：在实验观测过程中，有些物理量因太小而不能直接观测时，就要借助于声、光或叠加等方法，将待测量放大后再去观测，如游标卡尺、放大镜、望远镜、显微镜等仪器的设计都源于放大的思想。一些微小量的测量，如测纸的厚度，测金属丝的直径，测一段头发的质量等，常采用叠加放大法。

(4) 比较法：物理实验常通过对一些物理现象或物理量的比较，来达到异中求同的实验目的，此即比较法。如等臂天平、电桥等都是依据比较的原理。一些物理实验，如自感现象的观察、光谱分析等也是运用比较法设计的。

(5) 再现法：再现即模拟自然现象发生的条件，在实验室中重现自然物理现象的方法。如用棱镜对光进行色散就可以模拟彩虹，从而达到研究彩虹的目的。

当然，实验的设计方法不限于上述几种。

### 2.2.5 物理实验的地位和作用

物理学是一门以实验为研究基础的科学，物理规律的发现和物理理论的形成都是以物理实验为基础的，并且都必须接受实验的检验。物理学中的许多重大发现，都是通过复杂的科学实验实现的。实验为物理学提供了丰富的新事实和新规律，它既是物理学的基础，同时又是检验理论的唯一依据。伽利略的实验研究，特别是他将实验方法引进物理学，使物理学的研究由空谈变为实际，才使物理学的研究开始步入科学轨道。

**1. 实验是物理学理论基础**

物理学的特点是理论与实验并重，实验是研究物理学的基础。理论与实验脱节，将会使物理学的研究朝幻想方面发展。纵观物理学发展史，每一个物理理论的建立都是以物理实验为基础的。物理学的发展从总体上看是按照“实验→理论→实验→理论→实验……”的方式进行的。理论必须从实践中来，又必须回到实践中去，只有这样，理论才能上升为真理。实验上的重大突破往往是一个新理论的产生或者是理论的不断完善的过程。

人们对电磁学的研究虽然很早就已经开始，但进展非常缓慢，这是由于人们将电现象和磁现象分开来研究的缘故。1820 年丹麦物理学家奥斯特从实验中发现了电流磁效应后，才使物理学家改变研究方法，把电、磁现象联系起来研究，使电磁学的发展走向正确的轨道。1831 年英国伟大的实验物理学家法拉第（Michael Faraday，1791—1867）从实验中发现了电磁感应现象，为麦克斯韦建立电磁理论打下了坚实的基础。

迈克耳孙-莫雷实验和光行差实验否定了以太的存在，从根本上动摇了经典物理的基础。爱因斯坦以此为基础，在洛伦兹、彭加勒等人工作的基础上，大胆地否定了经典时空观，提出了光速不变原理和相对性原理，建立了相对论时空观，从而使物理学产生了一次革命性的飞跃。

因发现电子而获得诺贝尔物理学奖的英国物理学家汤姆孙在研究原子结构时，假设原子带正电的部分像“流体”一样均匀分布在球型的原子体积内，而电子则嵌在球体的某些固定位置。但由于这个模型在解释原子发光和原子衰变等方面存在许多问题，使人难以相信。为了进一步揭示原子的内部结构，只有靠新的实验手段去变革原子，才能去认识它。1909 年英国著名的物理学家卢瑟福和他的合作者设计了一个实验，用天然放射性源中放出来的高速 $\alpha$ 粒子去轰击金属箔，在实验中他们发现：大部分粒子能够透过金属箔，或者偏转了一个很小的角度，但还有极小部分（约 1/8000）的粒子却偏转了一个很大的角度，有的甚至接近于 180°，即几乎是反弹回来。这一种大角散射现象完全不能被汤姆孙模型所解释，于是卢瑟福于 1911 年提出了原子的有核模型，即原子结构的行星模型。当然它只是人们认识原子结构的一个阶段，随着科学的发展，行星模型又被玻尔的半经典模型所取代。

**2. 实验上的发现为物理学开辟了新天地**

实验上的发现具有一定的偶然性，实验上的一个新的发现会对物理学产生巨大的影响，有可能成为一个新理论的开始。一个优秀物理学家要能够善于抓住实验中出现的一些现象，这样才有可能在物理学上开辟一个新的研究方向。

近代物理学的发展可以说是从 X 射线的发现开始的，发现 X 射线的是历史上第一个获得诺贝尔物理学奖的德国物理学家伦琴。伦琴的这一发现纯属偶然，1895 年伦琴在研究阴极射线激发玻璃壁而发生荧光时，偶然发现在一两米外的荧光屏上出现了闪光，这使伦琴十分惊异，他一连几个星期吃住在实验室，经反复实验，他确信放电管发出了一种不同于阴极射线的新射线。由于当时人们对这种射线不了解，故此伦琴用数学上的未知数 X 表示，X 射线的发现开辟了近代物理学的新篇章。

1908 年，荷兰物理学家昂尼斯（Kamerlingh Onnes，1853—1926）成功地获得了液化氦。氦在一个大气压下被液化时，其温度为 4.2K。1911 年昂尼斯用这一低温技术成果来研究水银导线的电阻随温度变化的规律。他测得样品在温度为 4.2K 时，电阻骤降为零。当时，所有的理论都无法圆满地解释金属导体这种非零温度下零电阻现象。昂尼斯的这一发现立即轰动全世界，为此在全世界掀起了超导研究的热潮。

**3. 实验是物理理论的检验标准**

实验是检验真理的唯一标准。任何理论都来源于实验，又必须接受实验的检验。

从 17 世纪就有惠更斯的波动学和牛顿的粒子学之争，直到 19 世纪初杨氏的双缝实验和菲涅尔的单缝实验才使人们认识到光是波动而不是粒子。后来在 19 世纪末相继发现了黑体辐射和光电效应等现象，不能用光的波动学理论和电磁学理论解释。最终由普朗克和

爱因斯坦用光的量子理论才圆满地解释了黑体辐射和光电效应等现象。至此,人们才认识到光不仅具有波动性,而且具有粒子性的波粒二象性特征。

1864年英国杰出的物理学家麦克斯韦在奥斯特、安培(André-Marie Ampère,1775—1836)、法拉第等人工作的基础上建立了电磁学理论,并预言了电磁波的存在。但由于缺乏实验依据,他的理论在当时并不被人们所重视,直到1888年德国实验物理学家赫兹从实验中发现了电磁波,无可辩驳地证明了电磁波理论的正确性,才使人们接受电磁学理论,由此开创了无线电技术的新时代。

1905年爱因斯坦为解释光电效应现象,提出了光量子假说,但并不被人们所接受。当时著名的美国物理学家密立根(Robert Andrews Millikan,1868—1953)花费了10年时间做"光电效应"实验,目的是想用实验来否定爱因斯坦的光电子理论,但在精确的实验结果面前,他终于信服了,于1915年宣布证实了爱因斯坦的光电效应公式,接着美国物理学家康普顿和中国物理学家吴有训合作,用实验证实了康普顿于1923年提出X射线的量子散射理论,再一次证明了光量子理论的正确性,这才使人们接受普朗克、爱因斯坦的光量子理论。

普朗克、爱因斯坦的光量子理论通过密立根、康普顿等人的实验证实后,人们认识到光具有波粒二象性。这引起了法国物理学家德布罗意的极大兴趣,他用类比法于1924年提出了微观粒子也具有波粒二象性的物质波理论。1927年美国物理学家戴维孙和革末用电子衍射实验证实了电子波的存在,表明"物质波理论的正确性",为量子力学发展起到奠基作用。

从以上看来,实验在物理学中占有非常重要的地位,可以说离开实验就没有物理学的发展和进步。诺贝尔物理学奖从1901年到1998年共颁奖92次,总计有105个项目,145人获得诺贝尔物理学奖,其中实验项目(或人数)就占获奖项目(或人数)的三分之二以上。再从获奖时间上看,在实验上取得成果到获得诺贝尔物理学奖的时间较短,而从理论上取得成果到获得诺贝尔物理学奖的时间相对来说较长。因此,物理实验在物理学中的作用是非常重要的。对于一个物理工作者,不仅要掌握物理学理论知识,更需要掌握一定的实验技能,这样才可能获得成功。

## 2.3 理想化方法

自然界发生的一切物理现象和物理过程,一般都是比较复杂的,影响它们的因素也是多种多样的。如果不分主次地考虑一切因素,不仅会增加研究的难度,而且也不能得出精确的结果,相反还会对简单物理图像的分析也难以下手。因此,为了研究物理现象的规律性,就需要把复杂的问题转化为理想的问题,这样就在物理研究中产生了理想化方法。

理想化方法主要包括建立理想模型与设计理想实验。本节将先介绍理想化方法以及它的形式,通过几个例子讨论理想模型的特点和应用,结合大学物理中的多个理想实验讨论理想实验的实质、特点以及其在物理教学中的功能。

### 2.3.1 理想化方法的概述

**1. 什么是理想化方法**

无论是探索物理现象,揭示物理过程的规律,还是解决实际的物理问题,都需要建立理

想模型或理想过程。理想化方法就是借助于逻辑思维和想象力，有意识地突出研究对象的主要因素，完全排除次要因素和无关因素的干扰，在大脑中形成理想化的研究客体或相互联系，来探索物理世界内在奥秘的方法。它是一种科学抽象的方法。

理想方法在自然科学研究中占有重要地位，在科技史上多次被一些著名科学家运用，是一种重要的科学研究方法。中学物理乃至大学物理所研究的对象，严格说来，大都是理想模型，研究这些模型所遵循的规律，可以为解决大量复杂的物理问题奠定基础。因此，理想化方法不仅是物理学的研究方法，也是大、中学生学习物理学时必须掌握的一种重要思维方法。

**2. 物理学中理想化方法的形式**

在物理学中理想化方法主要表现为以下三种形式。

(1) 理想模型

所谓理想模型，是指在原型(物理实体、物理系统、物理过程)的基础上，经过科学抽象而建立起来的一种研究客体。它忽略了原型的次要因素，集中突出了原型中起主导作用的因素，摒弃了次要矛盾，突出了主要矛盾，如力学中的质点、刚体、光滑面、弹簧振子、连续介质、理想流体；热学中的理想气体、孤立系统；电磁学中的点电荷、匀强电场、匀强磁场、纯电阻、纯电感、纯电容、无限长螺线管、理想变压器；光学中的点光源、光线与光的直线传播、薄透镜；近代物理中的绝对黑体等。

(2) 理想过程

所谓理想过程，是指在研究物体的运动过程中忽略掉次要因素，只保留主要因素，将物体运动与状态的过程理想化，如匀速直线运动、自由落体运动、抛体运动、简谐振动、简谐波、单摆、完全弹性碰撞、完全非弹性碰撞、准静态过程、气体状态变化的等温过程、等压过程、等容过程、绝热过程、卡诺循环等。

(3) 理想实验

所谓“理想实验”，又叫作“假想实验”“抽象的实验”“思想实验”或“思维实验”，它是人们在思想中塑造的理想过程，是一种逻辑推理的思维过程和理论研究的重要方法。

在自然科学的理论研究中，“理想实验”具有重要的作用。作为一种抽象思维的方法，“理想实验”可以使人们对实际的科学实验有更深刻的理解，可以进一步揭示出客观现象和过程之间内在的逻辑联系，并由此得出重要的结论。

例如，作为经典力学基础的惯性定律，就是“理想实验”的一个重要结论。这个结论是不能直接从实验中得出的。伽利略曾注意到，当一个球从一个斜面上滚下而又滚上第二个斜面时，球在第二个斜面上所达到的高度同它在第一个斜面上开始滚下时的高度几乎相等。伽利略断定高度上的这一微小差别是由于摩擦而产生的，如能将摩擦完全消除的话，高度将恰好相等。然后，他推想说，在完全没有摩擦的情况下，不管第二个斜面的倾斜度多么小，球在第二个斜面上总要达到相同的高度。最后，如果第二个斜面的倾斜度完全消除了，那么球从第一个斜面上滚下来之后，将以恒定的速度在无限长的平面上永远不停地运动下去。这个实验是无法实现的，因为永远也无法将摩擦完全消除掉。所以，这只是一个“理想实验”。但是，伽利略由此而得到的结论，却打破了自亚里士多德(Aristotle，前 384—前 322)以来1000 多年间关于受力运动的物体，当外力停止作用时便归于静止的陈旧观念，为近代力学的建立奠定了基础。后来，这个结论被牛顿总结为运动第一定律。

爱因斯坦在建立狭义相对论时，曾经做了关于同时性的相对性的一个“理想实验”。即

当两道闪电同时下击一条东西方向的铁路轨道时,对于站在两道闪电正中间的铁道旁边的一个观察者来说,这两道闪电是同时发生的。但是,对于乘坐一列由东向西以高速行进的火车正好经过第一个观察者对面的第二个观察者来说,这两道闪电并不是同时下击的。因为,第二个观察者是在行近西方的闪电而远离东方的闪电,西方的闪电到达他的眼里的时间要早一点。因此,在静止的观察者看来是同时发生的闪电,在运动中的观察者看来却是西方先亮,接着东方再亮。同时性的相对性这一概念的提出,是狭义相对论建立过程中的一个关键。

爱因斯坦在建立广义相对论时,做了自由下落的升降机的"理想实验"。他设想:在自由下落的升降机里,一个人从口袋中拿出一块手帕和一块表,让它们从手上掉下来,如果没有任何空气阻力或摩擦力,那么在他自己看来,这两个物体就停在他松开手的地方。因为,在他的坐标系中,引力场已经被屏蔽或排除了。但是,在升降机外面的观察者看来,则发现这两个物体以同样的加速度向地面落下。这个情况正揭露了引力质量和惯性质量相等。爱因斯坦又设想了另一种情况的"理想实验",即升降机不是自由下落,而是在一个不变的力的作用下垂直向上运动(即强化了升降机内部的引力场)。同时设想,有一束光穿过升降机一个侧面的窗口水平地射进升降机内,并在极短的时间之后射到对面的墙上。爱因斯坦根据光具有质量以及惯性质量和引力质量等效的事实,预言一束光在引力场中会由于引力的作用而弯曲,就如同以光速水平抛出的物体的路线会由于引力的作用而弯曲一样。爱因斯坦预言的光线在引力场中会弯曲这一广义相对论效应,已为后来的观测结果所证实。

量子论的建立也同"理想实验"密切相关。在量子力学中,海森伯用来推导测不准关系的所谓电子束的单缝衍射实验,也是一种"理想实验"。因为,中等速度的电子的波长数量级为 $10^{-8}$cm 左右,这跟原子之间的距离属于同一个数量级。因而,只要让电子束穿过原子之间的空隙,就会发生衍射。但是,要想制成能够使电子发生衍射的单缝,首先就必须做到把单缝周围的所有原子之间的空隙都给堵死。实际上这是做不到的。在实验中,人们只能做到电子的原子晶格衍射实验,而无法实现电子的单缝衍射实验。

"理想实验"在自然科学的理论研究中有着重要的作用,但是,"理想实验"的方法也有其局限性。"理想实验"只是一种逻辑推理的思维过程,它的作用只限于逻辑上的证明与反驳,而不能用来作为检验认识正确与否的标准。相反,由"理想实验"所得出的任何推论,都必须由观察或实验的结果来检验。

### 2.3.2 理想模型

**1. 理想模型的特点**

理想模型是经过科学抽象而建立起来的一种绝对理想形态,具有科学的推测性。它具有以下两个特点:

其一,理想模型是抽象性和形象性的统一。模型的建立过程是一个抽象过程,然而模型本身又具有直观、形象的特点。

其二,理想模型是科学性和假定性的辩证统一。理想模型不仅再现了过去已经感知过的直观形象,而且要以先前已经获得的科学知识为依据,经过判断、推理等一系列逻辑上的严格论证。所以,具有深刻的理论基础,即具有一定的科学性。理想模型来源于现实,又高于客观现实,是抽象思维的结果,所以,又具有一定的假设性,只有经过实验证实了以后才被认可,才有可能发展成为理论。

**2. 大学物理中基本的理想模型**

物理学的研究对象是一个错综复杂、五彩缤纷的物质世界，物理学家在探求真知的过程中所建立的理想模型涉及力、热、光、电、原子物理等各个领域。众多理想模型的建立启迪物理学家创立新的科学理论，激励人们向着自然的更深层次去探索。为了对各个理想模型有一个更深刻的认识，下面对大学物理中常出现的理想模型简述如下。

1）力学

（1）质点

研究物体的运动时，我们常常把可以忽略其大小和形状的物体当作质点。质点是力学中很重要的一个理想的物理模型。

在许多实际问题中，物体的形状和大小与所研究的问题无关或者所起的作用很小，

我们可以在尺度上把它看作一个几何点，而不必考虑它的形状和大小，它的质量可以认为就集中在这个点上，这种抽象化的模型，叫作质点。

在哪些情况下，可以把物体看作质点？

回答此问题，需从具体事例谈起。有一些运动，如火车行驶时车厢的运动、刨子刨木头的运动、抽屉从桌子拉出来的运动等，这些运动物体上各点在任意时刻运动的快慢和方向都相同，这种运动叫平动。平动的物体上任意一点的运动都能代表整体的运动，因而可以不考虑平动物体的整体的大小、形状，把它当作质点。转动的物体上各点的运动情况是不一样的，就不能用一个点代表整体的运动，因而转动的物体就不可当作质点。

在力学中，把物体当作质点是有条件的、相对的。地球虽大，研究其公转时，可把它当作质点；但研究其自转时，则不可以把它当作质点。子弹头虽小，若研究其沿一定曲线击中某处目标的运动时，可视为质点。可见，能否把物体当作质点处理，必须对具体问题进行具体分析。一般来说，当物体自身的尺度远小于所研究运动的有关距离，而问题又不涉及转动时，可以把物体当作质点。

（2）刚体

我们研究问题时经常将物体视作质点来讨论。但是当它涉及物体的转动、自转及桥梁的平衡问题时，其形状、大小及其在力的作用下的形变都起重要作用，则必须考虑。虽然这样使问题变得复杂，但是当其形变很小，可忽略不计时，可将研究对象看成“刚体”进行研究。刚体和质点一样，也是一种理想模型，在研究问题中，只有当物体的大小和形状变化可以忽略不计时，才可以把它当作刚体看待。

将研究对象视为哪种理想模型，视问题的性质而定。如悬挂的大钟，研究它的平衡问题，可把钟看作质点；如研究其摆动，则需要将其视为刚体。研究刚体力学时，也常将刚体看作不变质点系，并运用已知的质点系的运动规律去研究。

（3）弹性体

当不考虑物体的形状、大小及形变时，我们总把研究对象简化为“质点”或“刚体”。但是任何物体在力的作用下都会发生形变，而且在某些问题中讨论其在力作用下形变的规律，也是很重要的，如斜拉桥上各悬索在力作用下伸长量的限度等。

若物体所受外力撤销后，在力作用下所发生的形变和体积的变化能够消失，则这种形变叫“弹性形变”，这种物体叫“弹性体”。“弹性体”也是一种理想模型，不存在绝对的弹性体，但房屋、地基、水库、堤坝、大建筑的薄壳等在形变极小时，在讨论一些问题时可看作弹性体。

2）热学

（1）孤立系

热力学研究的对象是由大量微观粒子组成的宏观物质系统。若系统和外界无任何相互作用，则称其为孤立系。所谓无任何相互作用是指与外界既无能量交换又无物质交换。当然，绝对意义下的孤立系实际上是不存在的。这样的系统甚至不可能成为研究对象。实际上，当系统与外界的相互作用十分微弱，其相互作用能量远小于系统本身的能量，在讨论中可以忽略不计时，我们就把系统看作孤立系。所以，孤立系的概念是一个理想的极限概念。

（2）平衡态

平衡态是宏观状态的一种较重要的特殊情况。系统不受外界影响的条件下，宏观性质（比如温度、压强、密度等）不随时间变化的状态叫作平衡态。这里所说的不受外界影响是指外界对系统不做功，不传热。当然，在实际中并不存在完全不受外界影响，而且宏观性质绝对保持不变的系统，所以平衡态只是一个理想的概念，它是在一定条件下对实际情况的概括和抽象。

虽然，平衡态是指系统的宏观性质不随时间变化，但从微观看，在平衡态下，组成系统的分子仍在不停地运动着，只不过分子运动的平均效果不随时间改变，而这种平均效果的不变，在宏观上就表现为系统达到了平衡态。因此，其平衡是一种动的平衡，即热动平衡。

（3）理想气体

在温度不太高、压强不太大的情况下，一定质量的气体基本遵守玻意耳定律、查理定律和盖·吕萨克定律。以这三条实验定律为基础，进一步推导出状态方程

$$\frac{pV}{T}=\text{恒量} \tag{2-3-1}$$

显然，也只有在这种条件下，不论什么气体方程都近似地符合状态方程。为了研究问题的方便，我们可以设想出一种新的气体，能够在任何温度和压强下都遵守状态方程，我们把能够严格遵守状态方程的气体叫作理想气体。

当然，理想气体是不存在的，它只是实际气体在温度不太高和压强不太大的条件下的近似。它是一种重要的理论模型，它是实际气体在压强趋于零时的极限。在通常压强下，可以近似地用这个模型来概括实际气体，压强越低，这种概括的精度越高。

从微观的角度来看，理想气体是忽略了气体分子之间相互作用的一个理论模型。当压强足够低时，气体足够稀薄，分子之间的平均距离足够大，其平均相互作用能量将远小于分子的平均动能，可以忽略；忽略分子的大小，将每个分子简化为质点；分子之间相互作用仅限于完全弹性碰撞。

3）光学

（1）细光束和光线

光束是指一个有限的立体角的区域内传播的光。当光束的横截面的限度比它的传播距离小得多时，这样的光束称为细光束（又叫物理光束）。光线是物理光束的进一步抽象，是指沿着细光束的轴线方向所作的几何线，它是表示光的传播方向的几何线。

（2）点光源

光源是指自行发光的物体。所谓自行发光是指其他物质形态与光这一物质形态的转换，例如电灯能把电能转化为光能，蜡烛能把化学能转化为光能，等等。这里所说的电灯、燃烧的蜡烛都是光源。

点光源也是一个理想的物理模型。当光源自身的限度远远小于光到物体之间的距离

时，而且光是向各个方向均匀发散的，这样的光源叫点光源。显然，点光源不一定是很小的光源。在几何光学中，由于不考虑光能的问题，故可将点光源看成几何点。这样的光源实际上是不存在的。但这样处理问题却使光路图简单明了，误差也很小。

（3）黑体

各种物体都有其不同的结构，对外来辐射的吸收以及它本身对外的辐射都不相同。但是有一类物体其表面不反射光，它们在任何温度下吸收射来的一切电磁辐射，这类物体就叫绝对黑体，简称黑体。

黑体只是一种理想模型。做一个闭合的空腔，在空腔表面上开一个小孔，从外面射来的辐射，经过小孔射入空腔，要在腔壁上经过多次反射，才能有机会射出小孔。因此，在多次反射过程中，外面射来的辐射几乎全部被腔壁吸收。这样，此空腔小孔的表面就可以模拟黑体表面。

4）电学

（1）点电荷

实验表明，两个静止的带电体之间的作用力（静电力）除与电量及相对位置有关，还依赖于带电体的大小、形状、电荷的分布情况。如果带电体的线度比带电体之间的距离小得多，那么静电力就基本上只取决于它们的电量和距离，满足此条件的带电体叫作点电荷。

点电荷的概念类似于力学中的质点的概念。带电体能否被看作点电荷，不仅取决于本身的大小，而且取决于它们之间的距离。例如，两个半径为 1cm 的带电球，当球心距离为 100m 时，可以充分精确地看作点电荷；当球心距离为 3cm 时再看作点电荷就会带来很大误差。但是，究竟带电体的线度比距离小多少才能被看作点电荷，却没有一个绝对的标准，它取决于讨论问题时所需要的精确程度。带电体一旦被看作点电荷，就可用一个几何点标志它的位置，两个点电荷的距离就是标志它们的位置的两个几何点之间的距离。

（2）电路模型

实际电路都是由一些按照需要起不同作用的实际元件或器件组成的，例如变压器、电池、晶体管以及各种电阻器和电容器等，它们的电磁性质较为复杂。最简单的例如一个白炽灯，它除具有消耗电能的性质（电阻性）外，当通有电流时还会产生磁场，就是它还具有电磁性。但电感微小，可忽略不计，于是可以认为白炽灯是一电阻元件。

为了便于对实际电路进行分析和用数学描述，将实际元件理想化（或称模型化），即在一定条件下突出其主要的电磁性质，忽略其次要因素，把它近似地看作理想电路元件。由一些理想电路元件所组成的电路，就是实际电路的电路模型，它是对实际电路电磁性质的科学抽象和概括。在理想电路元件中，主要有电阻元件、电感元件、电容元件和电源元件等。这些元件分别由相应参数来表征。

5）原子物理学

汤姆孙最早提出了一个原子的结构模型，即“枣糕模型”。他设想原子的带正电部分是一个原子那么大的、冻胶状的球，正电荷均匀地分布着，在这球内或球上，有负电子嵌着。这一模型解释了原子呈中性和原子发光现象。但随着粒子散射实验，卢瑟福提出了“行星模型”，建立了原子的核式结构。即原子是由居于原子中心带正电的原子核和核外电子构成的，原子的带电部分集中在很小的体积中，但它占有绝大部分的质量。鉴于“行星模型”与经典电磁场理论矛盾，丹麦物理学家玻尔提出氢原子模型。量子力学建立后，人们认为原子核周围存在电子云，电子没有确定的轨道，玻尔模型中的轨道是电子出现概率最大的地方。

关于原子模型的演变过程充分地说明了理想模型是变化的、发展的。它也说明了人们认识客观世界是由浅入深、由片面到全面、由局部到整体逐步深入的，物理模型的建立也是逐步深入的。

大量的正确的模型的成功，体现了建立模型应用模型方法的意义和作用。不过，也有不少模型在科学面前最终被证明是错误的，因而被人们抛弃了。但是，它们从反面却给人们以深刻的启示，在人类认识史上为建立新的模型仍然起到了一定的积极作用。

**3. 用理想化模型时应注意的几个问题**

物理教科书中所研究的对象都是抽象成一定的理想化模型，如力学中的质点、刚体，热学中的理想气体，光学中的点光源、光线等，都是我们非常熟悉的理想模型。但是在具体的问题中怎样用理想模型来代替实际物体呢？

有时候我们常会遇到这样一些情况，即所研究的某些物体与某些理想化模型很相似，例如，电子、质点、中子等基本粒子很像“质点”；许多坚硬物体很像“刚体”；橡皮绳、弹簧、皮球等物体很像“弹性小球”等。类似这种情况，我们必须明确，究竟用什么模型来代替实际物体，决不可以它们的外形特征或某些表面因素为依据，而应该以研究问题的实际需要不同抽象成为不同的理想模型。如对同一个皮球，如果需要研究它的位置移动情况或者与别的物体发生碰撞前后的速度变化情况，则可将皮球视为“质点”；如果要研究皮球的转动，则可将皮球视作“刚体”；如果要研究皮球的弹性形变，则可将皮球视为“弹性体”。总之。理想模型和实际物体是两回事，前者是科学抽象，后者是客观存在，不可混为一谈。

有些问题的研究往往需要选用两种或两种以上的理想化模型。例如：我们研究小电珠或烛焰在远离玻璃透镜时的成像情况，可选取三种模型来研究，即小电珠或烛焰视为“点光源”，把实际透镜视为没有厚度的“薄透镜”，把光源发出的光线视为“沿直线传播”。又如，弹簧振子本身是一种理想模型，同时它又可看作“质点(小球)”和“弹性体(弹簧)”这两种理想模型的组合。

总之，在研究实际问题时，我们应该根据实际需要和具体要求，合理地利用理想模型来代替实物，使问题得到简化，并得到解决。

**4. 理想模型在自然科学研究中的重要作用**

(1) 具有对研究对象极度简化和纯化的作用

物理模型最明显的作用是对所研究的物理问题的处理极度的简化和纯化。从模型推出的结果与客观实际又不会发生大的偏差，即在一定的场合、一定条件下，作为一种近似，可以把实际事物当理想模型来处理，可以将模型的研究结果直接用于实际事物。

(2) 理想模型法可以帮助人们认识事物的规律性

人们通过对理想模型的研究，可以达到对实际事物或现象的性质的规律认识，对于复杂的物理学问题，先研究它的物理模型，然后，将其研究结果加以修正，以便与实际对象的本质相符合。例如，由理想气体状态方程 $pV=nRT$ 得出的结论与实际气体不相符合，这是因为理想气体不考虑气体分子的大小和分子间的相互作用。但是，实际上分子是有大小的，为了弥补这个不足，必须从理想气体体积 $V$ 中减去分子的体积 $b$；实际上分子间的相互作用也是不能完全忽略的，所以，测定的压强值 $p$ 必须加上一个反映气体引力的压强修正值 $p'=\frac{a}{V^2}$，则得出实际气体状态方程，即范德瓦尔斯方程：

$$\left(p+\frac{a}{V^2}\right)(V-b')=kT \tag{2-3-2}$$

（3）物理模型法可以促进人们想象力和逻辑思维能力的发挥和发展

物理模型是经过抽象思维构思出来的，它可以超越现有的条件，指示研究方向，有助于提出科学预见和科学假设。例如，电磁波的预见是麦克斯韦在他建立的电磁场模型的基础上由麦克斯韦方程组得出的结果。

总之，对于微观世界和宏观世界乃至宇宙的研究，都必须借助于物理理想模型进行。

**5. 理想模型的局限性**

建立一定的物理模型，对于物理学理论的建立和发展发挥着重要的作用，但人们常常被许多理想模型的功绩所折服，而忽略了问题的另一方面，即在物理学理论发生重要突破的关键时刻，一些崭新的概念或原理往往是挣脱原有种种模型的束缚，且不倚赖于任何具体模型而诞生出来的。

以原子结构的玻尔模型为例。为了解释当时已经发现的原子结构的规律性，玻尔大胆地越出了经典物理学原理的限制，提出了关于电子轨道量子化的假定，从而取得重要的成果。但这一模型并没有完全挣脱经典物理学的束缚，表现为它不能解释电子运动的稳定性，也不能解释处在不同状态之间的跃迁概率问题，这些问题在量子力学出现之后才得到解决，测不准原理告诉人们，像电子这样的微观粒子本来就不可能有经典的运动轨道。这就是说，玻尔模型是新的量子力学普遍原理的束缚。

### 2.3.3 理想实验

理想实验，作为理想化方法的一种重要表现形式，它在实验仪器、实验条件以及实验过程中实现理想化，是进行理论研究时必不可少的一种手段。物理学发展过程中，理想实验在发现新规律、反驳谬论、建立新理论等方面均有许多成功的运用。

**1. 理想实验的特点及作用**

理想实验，也称为假想实验、思想实验或思维实验，它不是真实的实验，它是人们在头脑中按实验特点塑造的理想过程，是一种逻辑推理的思维过程。

（1）理想实验的特点

理想实验虽然也叫实验，可它同真实的科学实验是有原则区别的，真实的科学实验是一种实践活动，而理想实验则是一种思维的活动。因此，理想实验有一些真实实验所没有的突出特点，主要表现为：

理想实验由于思想上可取理想化条件，所以没有实验误差。

理想实验中有超越当时技术水平的实验仪器或实验场景，这种思想上设计的“仪器”一旦条件具备，理想实验就可能成为现实。例如“自由落体实验”中“没有空气阻力”的条件，后来在真空管中实现，证实了伽利略的推论：重物与轻物下落一样快。但较多的理想实验也许根本不能成为现实，例如“惯性实验”中“运动方向上不受力”的条件难以实现，只能靠实验间接检验。

理想实验十分突出主要因素，所以易于找到主要规律。理想实验是在真实科学实验基础上，抓住主要矛盾，忽略次要矛盾，再以一定的经过长期实践活动总结并证实了的逻辑法则为根据，对实际过程做出更深入一层的抽象推理分析，以便进一步揭示出客观现象的本质面目，找到主要规律。

（2）理想实验的作用

理想实验是自然科学理论研究中的一种重要方法，在科学研究中具有重要作用。一方面，理想实验可以用来推翻不合理的理论，例如，伽利略用理想实验否定了亚里士多德“重物自由

下落较轻物快”的错误观点；另一方面，理想实验还可用来建立新理论，例如，伽利略的惯性实验，它为牛顿第一定律的建立铺平了道路，成为牛顿第一定律的基础；当然，理想实验还可用在反驳等方面。总之，作为一种抽象思维的方法，理想实验可以使人们对实际的科学实验有更深刻的理解，可以进一步揭示出客观现象和过程之间内在的逻辑联系，并由此得出重要结论。

**2. 理想实验的局限性**

还应指出的是，理想实验在它未能转化为真实实验之前，不能作为检验真理的标准，它仍需受到实践的检验。因为，它实质上不是实践，而是思维活动的一种形式，还含有违反新的客观实际的危险性，还可能导致错误的结论。

例如，伽利略从惯性实验出发，又进一步推理，得出“惯性运动必须是圆周运动”的错误结论。他认为：惯性运动不能是匀速直线运动。因为若地面无摩擦，物体将沿地面惯性运动下去，而地球表面是圆的，所以物体将围绕地球表面作一圆周运动。在这里，伽利略犯的错误是：惯性运动的前提是不受任何力，但他只注意到了在运动方向上不受任何力，而忽视了圆周运动中物体所受合力不等于零，因此圆周运动就不可能是惯性运动的事实。后来，笛卡儿将其修正为直线运动，这之后，“惯性实验”才真正成为牛顿第一定律的“实验”基础。

## 2.4 比较与分类

比较与分类是人类认识事物基本的逻辑方法，可以应用在自然科学研究、社会科学研究以及教育科学研究当中。人们认识事物开始于比较，而为了使认识系统化、深刻化，必须进行分类，从而再结合其方法进行更深入的研究，物理学的研究和学习也不例外。

### 2.4.1 比较

**1. 比较及其客观基础**

比较是确定研究对象之间差异点和共同点的思维过程和方法。

各种物理现象和过程都可以通过比较确定它们的差异点和共同点。比较是抽象与概括的前提，因此是物理学研究中经常用的主要方法。客观世界既是统一的，又是千姿百态的，这是比较法产生的客观基础。物理运动形态的种类无论怎样繁杂，怎样千变万化，它们总是相互区别又相互联系的。无论在空间上同时并存的事物之间，还是在时间上先后相随的事物之间，都可以找到它们赖以相互区别的特点，这就是事物的特殊性。因为一切物理事物及其运动都是相互联系相互依存的，所以也可以找到它们之间相同的地方。事物的同一性寓于事物的特殊性之中。这就是说，任何事物之间无论有多大的差异点，也总会存在共同点；反之，无论有多少共同点，也总会存在着差异点。这是由物质运动的同一性和多样性所决定的，即物质运动的同一性和多样性是比较方法的客观基础。

例如，物理学、化学和生物学，就其研究内容来说有着明显的区别，这就体现了它们的差异性。但是这种差异不是绝对的，现代科学中的新兴学科物理化学、物理生物等就体现了它们之间的联系。这种联系由它们的同一性所决定。

在物理学内部更是如此，动量和动能，它们都与物体的质量和速度有关，都是描述物体机械运动的量，这是它们的共性所在。但是，它们又存在着质的不同，动量是矢量，是用机械运动来描述机械运动变化的量，它只能从一个物体转移到另一个物体，在传递过程中不能变成其他运动形式。而动能是标量，它既可以在相互作用的物体之间相互传递，也可以变成其他形式的能——势能和热能。这就是同中有异，异中有同，不存在只有差异而没有同一的物

理现象和过程，也不存在只有同一而没有差异的物理现象和过程。

在比较过程中，条件是十分重要的。所谓同一和差异都是在一定条件下和一定范围内的同一和差异。在特定的范围内，一定的条件下相比较的物理现象和过程可能是互有相关的，但是在另外的范围和条件下可能是密不可分的。

例如，物体的质量和速度，在宏观低速的经典力学中是互相独立互不相关的，而在相对论中则遵从关系式 $m=\dfrac{m_0}{\sqrt{1-v^2/c^2}}$，这表明在经典力学中所谓的质量不变性是相对的，在高速的情况下，质量和速度是密切联系在一起的，由速度变化而引起的质量变化是不能忽略的。时间和空间也是如此，在牛顿力学体系中它们是相互独立的，似乎没有同一性，但在相对论中，也成为有密切联系的事物，这种联系就是它们的同一性所致。总之，同一和差异是不可分割地联系在一起的。事物及其运动变化中的这些特性、内在联系就是比较法的基础。

**2. 比较的类型**

在物理学研究领域中，由于研究对象的广泛性和多样性，进行比较研究的形式也是多种多样的。有的是比较同一物理现象在一定时间内前后的变化情况；有的是比较研究对象随条件变化的情况；有的是比较几个不同物理运动的情况；有的甚至要将物理领域中运动和其他科学领域中的运动情况相比较。这些比较，有的要在静态下进行，有的要在动态中进行。而且根据研究工作的需要，比较的手段也各不相同，所以，比较的类型是多种多样的，比较的分类也只能是相对的。

一般地讲，比较有三种类型。

其一，异中求同的比较，即指比较两个或两个以上的对象而找出其相同点。比较的模式是：

| 对象 | 被比较的特性 |
|---|---|
| A | a、b、c… |
| B | a、b、c… |

所以，A 与 B 两对象有相同的特性 a、b、c…

其二，同中求异的比较，即指比较两个或两个以上相似的对象，找出其相异点，比较的模式是：

| 对象 | 被比较的特性 |
|---|---|
| A | a、b、c… |
| B | a′、b′、c′… |

所以，A 对象以特性 a、b、c…与 B 对象相异

其三，同异综合的比较，即指比较两个或两个以上对象的相同、相异点，比较的模式是：

| 对象 | 被比较的特性 |
|---|---|
| A | a、b…p、q… |
| B | a、b…p′、q′… |

所以 A 对象以特性 a、b…相同于 B 对象，以特性 p、q 相异于 B 对象

**3. 比较的作用**

任何一项物理研究不管是自觉还是不自觉都包含着比较的成分。在物理学众多领域的实验和理论研究中,比较方法自古以来都占有一定的地位,发挥着重要的作用。

(1) 比较是建立物理概念、发现物理规律的基本方法

物理概念是物理现象和物理过程的本质属性在人头脑中的反映。它是科学思维的产物。而科学思维的第一步就是要把研究的众多物理现象和物理过程的性质进行比较,找出它们的同一性和差异性,并利用分析、综合、抽象、归纳等思维方法,抽取它们的本质属性,进而形成物理概念。例如,我们观察汽车在公路上行驶,飞机在天空飞行,木块沿着斜面下滑,机器运转,钟摆摆动,雨滴下落等,对这些物体的运动进行比较、分析和综合后,我们才认识到它们的表现形式虽然不同,但都有一个共同点,即一个物体相对于另一个物体的位置发生了变化,当抛开这些物体运动的具体形象,而抽象出它们的共同特征——相对于参照物位置的变化时,就形成了机械运动的概念。

物理规律是物理现象和物理过程在一定条件下发生和变化的必然趋势以及内在本质联系的反映。人们认识它们以前,它们已经客观地存在着。但是,它只能寓于特殊性之中。人们只有在研究大量的个别现象和过程并进行比较后,才能准确、全面地概括和抽象出它们的特性和规律,发现它们的同一性。例如,电磁感应定律,是法拉第等人对大量实验事实进行研究、比较之后才提出的。

(2) 比较是物理学研究中进行鉴别和测量的重要方法。利用木星卫星的掩蚀测光速就是一个很重要的例子

对光本质的认识是一个漫长的历史过程,其中,光的传播是否需要时间是关键的问题之一。开普勒认为光的传播是瞬时进行的,不需要时间,笛卡儿也持同样的观点,他认为光是压力的传递过程。伽利略认为光即使传播得很快,也是需要时间的,后来他想利用木星的卫星频繁掩蚀的现象进行光速的测量。这一工作是由丹麦天文学家罗默等人实现的。在这个系统里,木星的卫星蚀一方面作为一个信号供地球上人来观察,另一方面此卫星蚀的周期过程又是一个准确的时钟。如果地球相对于木星的距离不变,或者光速为无限大,则两次蚀之间的时间间隔是一定的。但是,众所周知,光速不是无限大,并且地球每时都在改变着它与木星的距离,所以在地球上看到的木星的卫星相邻蚀之间的时间间隔是变化的。显然这个变化与地球相对于木星的距离的变化和光速不为零有关。

**4. 比较的局限性**

比较法虽然是物理学研究中的重要思维方法之一,但是它的作用也仅仅是在比较鉴别之中。而且它与其他思维方法一样,必须要有观察和实验作为基础。同时,运用比较法,要受到人的洞察能力、理论水平、指导思想、实验设备等条件的制约。对同样的材料,由于各人的观点不同,看问题的角度不同,会出现"横看成岭侧成峰"的情况,使比较得出相反的结果。这就是说,比较的结果具有或然性,正确与否要由实践来检验。

## 2.4.2 分类

某些不同的物理现象具有某种同一性,在物理学研究中常常以某种同一性作为标准,将对象归合为一类,在这一类中,又根据对象的差异性划分为较少的类。于是将物理现象区分为具有一定从属关系的不同等级的系统,这就是物理研究中的分类方法。

分类要以比较为前提,因为没有比较就不可能确定研究对象的共同点和差异点,也就无法进行分类。分类与比较一样,都是以物理现象和物理过程本身固有的区别和联系作为自

己的客观基础的。由于客观的物理事物有多方面的属性，各事物间有多方面的联系，因而人们都可以根据不同的实际需要，依据不同的分类标准，对物理事物进行各种分类，从而得到不同形式的系统网络，如图 2-4-1 所示。

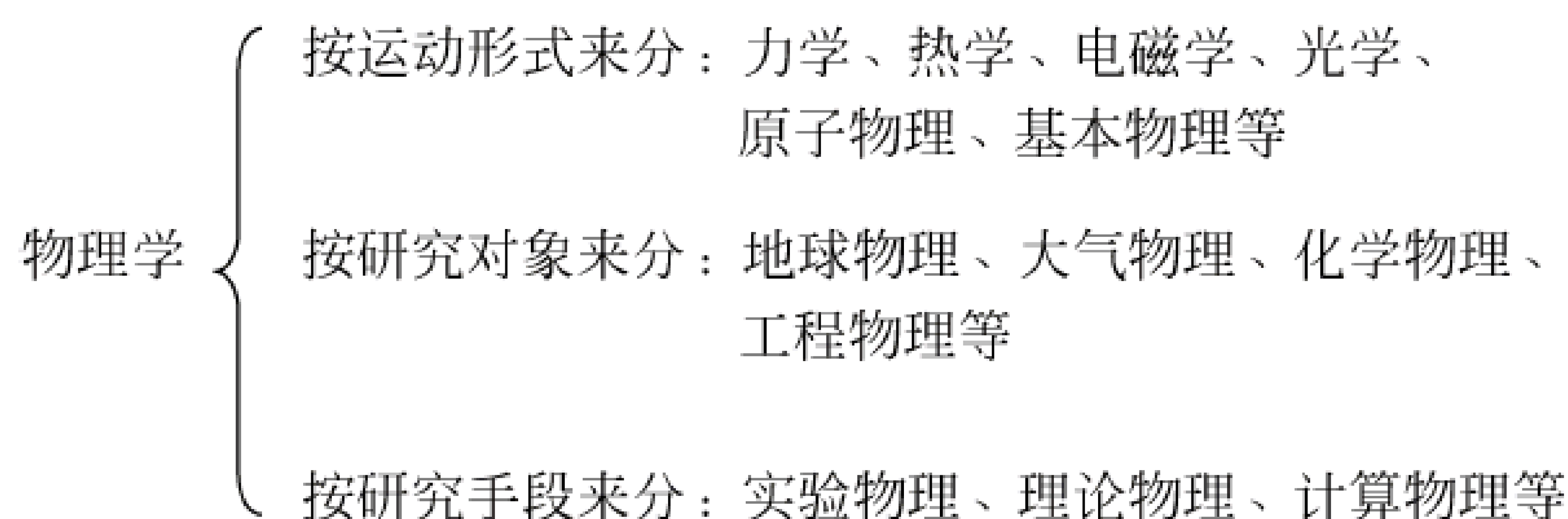



图 2-4-1　物理学学科的分类

**1. 分类的原则**

尽管可以根据实践需要，依据不同的分类标准进行不同的分类，但物理学的分类绝不是研究者可以随心所欲进行的，而是要依据物理事物本身的特点、运动规律以及物理事物之间的差异和内在联系进行分类。

(1) 分类要反映物理事物及其运动的内部次序。这就是说，研究者在对物理对象着手分类以前，必须把要分类的物理事物的本质属性进行比较和鉴别，弄清它们的同一性和差异性。而事物的同一性和差异性又是分层次的。这就需要正确地区分“种”和“类”(关于“种”和“类”的含义，请读者参阅本书 1.1 节)，搞清它们的从属关系，以形成一个符合自然界本来面目的系统。当然，这种从属关系的确立是与人们对事物的认识水平有密切关系的，它有一个历史发展过程。因此，物理学的分类演变也必然反映了物理学的发展过程。

(2) 遵循分类的穷尽原则。从逻辑上说，分类就是把一个种概念划分为若干属概念。种概念叫母项，属概念叫子项。穷尽性的意思就是划分出来的子项的外延之和，必须等于母项的外延。例如，我们根据导电性来对材料进行分类，如果分成导体和绝缘体，就没有包括所有的材料，因为还有半导体。

(3) 遵循分类的排他性原则。为了保证分类的排他性，必须注意以下两方面。分类的层次性，即分类要逐层进行。而且子项保持在同一层次，各层次的子项不能混淆。例如，把运动划分为直线运动、曲线运动、斜抛运动，就混淆了不同层次。另外，母项与子项要在相邻两个层次，不要跨越太多层次。例如，把运动分为直线运动、匀变速运动、自由落体运动，这样就跨越了几个层次。

**2. 分类的作用**

首先，分类为研究工作指明方向，促进物理学发展。

虽然物理学的分类是以比较为前提的，但比较与分类又是交替进行的。在初步分类之后，只有将所比较的物理现象和过程进行有效的分类，使它们有一定的条理性和系统性，才能进行更准确的实质性比较，找出事物间的本质区别和联系，进而探讨各类事物之间的转化关系。这样就为科学研究提供了更明确的线索，从而促进物理学研究的进展。物理学发展的初期就显示了分类的这种作用。如伽利略以前的力学把运动不恰当地分为“自然运动”和“强迫运动”，使这一时期对机械运动的研究处于思维混乱的状态之中。伽利略经过比较分析认为这种分类方法是错误的。他认为应该根据运动的基本特征量——“速度”对运动进行分类，由此提出了匀速运动和变速运动的新分类方法，从而使运动理论的研究取得重大进展。

其次,分类整理后的物理理论便于传授和检索使用。只有对物理知识进行分类整理,才能使其条理化和系统化,使各种物理现象、物理过程的特征和规律井然有序,由浅入深,由简到繁,由表面到本质,由低级到高级,以便物理知识的传授和学习能够做到循序渐进,逐步深入。

**3. 分类的局限性**

物理学中的分类同其他学科的分类一样,是相对的。各类别之间虽然有明确的区别,但是没有固定不变的界线和不可逾越的鸿沟,不管区别多大,总有这样或那样的联系。因此,不应将分类绝对化。特别是在具体的研究中不应该被已有的类比所禁锢。否则,就会延误科学的发展。在19世纪,发现电火花对化学分解和化合有明显的影响,这本来就是一个值得认真研究的课题。但是,由于人们墨守既成的分类,物理学家宣称,这是化学的事情,而化学家宣称,这是物理学的事情,双方都称与己无关,不去研究,结果延误了电化学这门学科的产生和发展。

## 2.5 类比与模拟方法

类比方法就是比较事物的相同点,模拟方法是根据研究对象的特点,先设计出与被研究对象或过程(即原型)相似的模型,然后通过模型,间接地研究原型规律性的逻辑方法。在物理学的发展历史上,有许多通过类比方法获得重大突破的例子。弹簧振子的简谐振动类比电磁振荡、机械波类比波动光学是当前大学物理学通用的方法。模拟物理实验几乎是现当代物理学研究的主要方法。

### 2.5.1 类比

在物理学的研究中,常常用已知的现象和过程(包括物理学以外的现象和过程)同未知的物理现象和过程相比较,找出它们的共同点、相似点或相联系的地方。然后,以此为根据推测未知的物理现象和过程也可能具有已知的现象和过程的某些特性和规律。这就是物理学研究中的类比方法。其模式如图2-5-1所示。

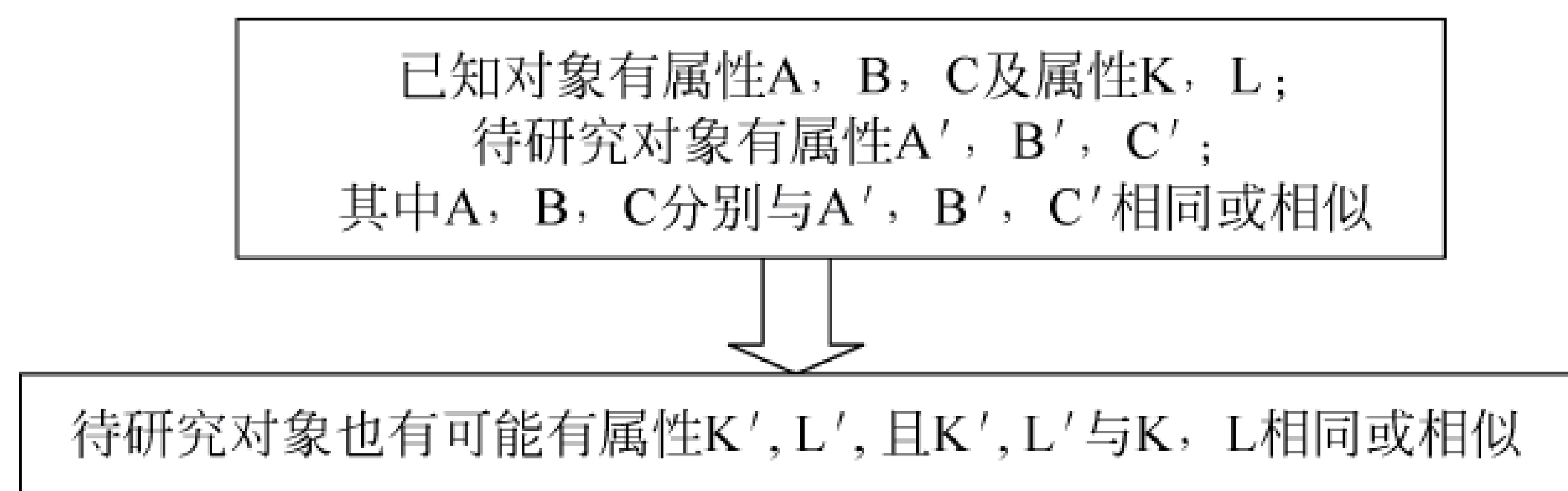



图2-5-1 类比法的模式

运用类比方法能启发和开拓我们的思维,能给我们提供解决问题的线索,因而在物理学中得到了广泛的使用。

**1. 类比法在物理学研究中的作用**

可以说,类比法贯穿于物理学的全部发展过程,不论是在经典物理的形成过程中,还是在近代物理的发展过程中,乃至对高新技术的提出,类比法都起到过不容忽视的作用。总结起来,主要有以下几个方面的作用。

(1) 类比法能够明确物理学研究的方向和路线,使研究工作少走弯路

从物理学的发展史来看,当研究陷入困境时,往往都是通过类比法去打开一片新的天地。使人们在“山重水复疑无路”的时候,看到“柳暗花明又一村”。20世纪30年代中期,人

们已经知道原子核是由质子和中子组成的。原子核一般都很稳定,这表明质子和中子结合得很紧,人们称这种结合力为核力。那么,核力是怎样产生的呢?这个问题长时间地困扰着物理学家。在原子核结构研究中,这是一个很重要的、迫切需要解决的问题。1935年日本物理学家汤川秀树(1907—1981)将核力和电磁力相类比,提出质子和中子是通过一种核力场结合在一起的。由一方放出粒子,另一方吸收粒子而相互作用。他称这种粒子为介子,在此基础上他还估算出了介子的质量。物理学家沿着汤川的方向和路线进行研究,1947年英国物理学家鲍威尔(Colin Luther Powell,1903—1969)终于找到了汤川的介子,证实了汤川的预言,使陷入困境的核力研究迎来了胜利的曙光。在今天看来,尽管汤川的预言对核力的认识还不够深入,但是他的方法论的意义却是深远的。在60年代以后,许多高能物理实验揭示了中子和质子也有自己的结构,是由称为层子的粒子组成的。层子之间是如何实现相互作用的呢?物理学家们还是像汤川那样采用类比法,设想一种称为"胶子"的东西。层子(国外学者称其为"夸克")通过放出和吸收胶子实现相互作用。胶子是否实际存在还需要实验验证,但是它启示了目前物理学研究的一个方向,不少物理学家正沿着这个方向艰难地探索着。

(2) 类比法是提出科学假说、做出科学预言的重要途径

在物理学研究中,运用类比法有助于科学研究者冲破学科和专业的界限,借他山之石,攻我之玉。因为进行类比的两个(或两类)研究对象,可以是同类的,也可以是不同类的;进行比较的属性和关系,可以是某一种,也可以是某几种;可以是本质的,也可以是非本质的。也就是说,类比的方法可以适用于同一运动形式的各种自然事物和现象,就是那些距离我们遥远的天体和微小的基本粒子都可以用类比的方法进行研究。物理学发展史上的许多科学假说和科学预言都是运用类比法而创立的。如卢瑟福把原子和太阳系类比提出了原子结构的太阳系模型;惠更斯把光和声类比提出了光的波动说;狄拉克对正电子的预言;德布罗意对物质波的预言;麦克斯韦将电磁感应定律与安培环路定律相比时,发现其中的不对称性,并由变化磁场能产生电场推断变化电场能产生磁场,提出了位移电流假设,建立了麦克斯韦积分方程组,从而从理论上预言了电磁波的存在。

(3) 应用类比法,可以获得重要的科学发现和科学发明

类比能够诱发出灵感。在物理研究中,经常会遇到这样的情况:研究者在沉思于某一课题,不得其解时,突然,由于某种信息的触发,通过类比,灵感突然而至。而这种灵感催生了很多重要的科学发现和科学发明。正如开普勒(Johannes Kepler,1571—1630)所说:"我珍惜类比胜于任何别的东西,它是我最可信赖的老师,它能揭示自然界的秘密……"牛顿把苹果和行星类比,激发了万有引力定律的产生;阿基米德(Archimedes,约前287—前212)把王冠和人体类比,造就了阿基米德定律的产生;当德布罗意关于物质波的论文于1924年发表以后,薛定谔受到了很大的启发,他沿着另一条思路将经典力学和几何光学做出如下类比:经典力学和几何光学具有完全相同的形式,几何光学又是波动光学的近似,因此,经典力学也可能是波动光学的近似。在此类推结果的引导下,薛定谔做了很多尝试,最终建立了波动力学的崭新体系,为现代物理学的发展做出了突出贡献;"多普勒效应"最初是关于声音传播的定律,多普勒把光和声进行类比,指出"多普勒效应"不仅适用于声波,也适用于光波,以后,哈勃等天文学家根据"多普勒效应"解释天文学上的"红移现象"取得了成功。

(4) 类比法可以更形象直观地揭示研究对象的特性和规律,使之更容易为人们理解

将抽象难懂的现象与过程和比较具体形象的现象与过程相类比,能够给人们提供一种

具体的模型,从而有利于人们理解。瑞典物理学家摄尔西斯等人发现不同的液体混合总体积减小的现象后,与下面情况相类比:两个容器分别装满铅球和铅粒,使铅球和铅粒混合后总装不满原来的两个容器,因为铅粒可藏于铅球的缝隙中。从而得出液体总体积减小的原因是一种液体分子跑到另一种液体分子空隙中去了。这个类比给人形象具体的模型,揭示了液体混合的一种特性。牛顿曾经认为光的折射和色散是不可分割地联系在一起的,因此,在研究望远镜时,为了避免色散现象,集中力量研究反射望远镜。后来,欧拉把透镜和人的眼睛作类比,眼睛是由角膜物质、水晶体、玻璃体等具有不同折射率的物质组成,视网膜上的像虽然是通过眼睛折射形成的,但是没有色散。因此,他认为折射和色散没有一种必然的联系。英国物理学家多伦多沿着欧拉的方向进行大量的研究,终于用两种折射率不同的玻璃适当配合做成透镜,避免了色散,使望远镜的研究进入了一个新的阶段。

**2. 物理学中类比法的主要形式**

自然界的事物形形色色,事物属性之间的关系也多种多样,导致类比的形式也多种多样,常见的主要有以下几种。

(1) 因果类比

因果类比是根据两个对象(或两类对象)的相似特性,推断其各自属性之间可能具有的相同因果关系而进行的类比推理。其模式如图 2-5-2 所示。

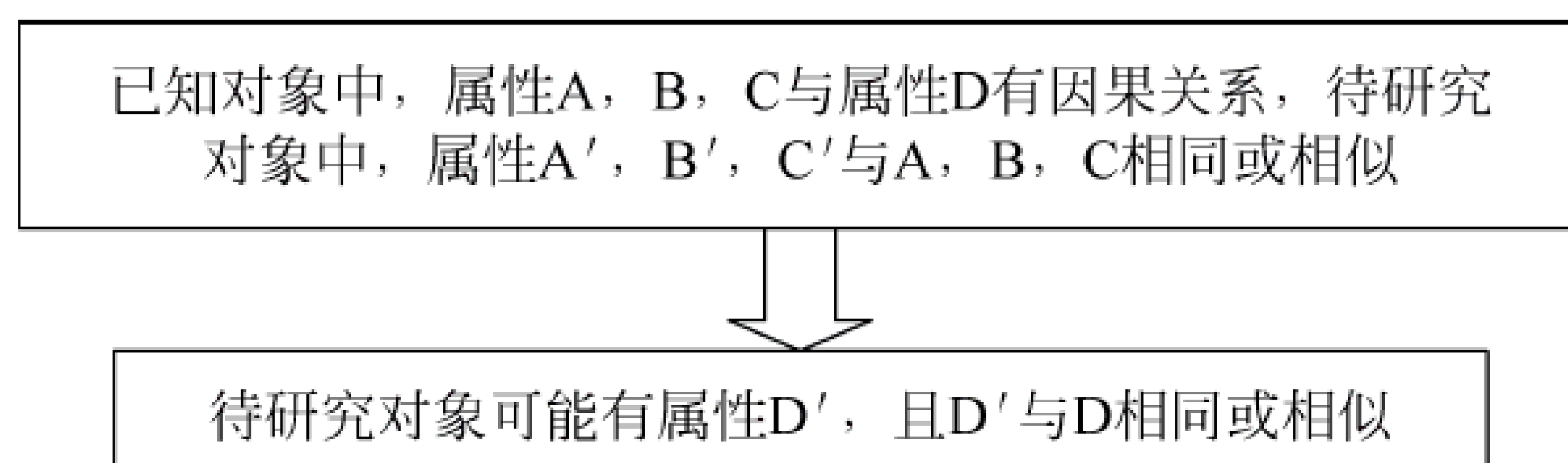



图 2-5-2　因果类比

这种因果类比的方法,在物理学史上曾经被牛顿所用,从而发现了万有引力定律。在我们的物理教学中也常常用到这种方法。例如:我们物理教材中电压与水压所做的类比就是因果类比,用连通器里的水流类比电路的电流,水泵类比电源,水压类比电路的电压,从而得出,电键闭合后,电路就有电流。电路里的自由电子之所以能够定向移动形成电流,是因为电源的正极有多余的正电荷,从而在连接电源两极的电路中产生了电压。电压是使自由电荷发生定向移动并形成电流的原因。

(2) 数学形式的类比

数学类比是根据两个(或两类)对象可能具有的属性之间的某种函数关系进行的类比推理。由于物理规律一般都可以用数学形式定量地抽象概括出来,数学形式相同的物理现象往往遵循一种更深刻、更本质的规律,这就决定了它们有更深刻的共同特征。因此数学形式的类比往往能使人们认识到研究对象在感性直观方面很难认识到的特性和规律。这种类比形式有两种模式。

① 由两个对象的主要属性相同或相似,推论出它们的数学形式相同或相似。其模式如图 2-5-3 所示。

比如,德布罗意在 1924 年提出物质波公式时,他的推理过程如图 2-5-4 所示。此关系式后来被电子衍射实验所证实。

再如,库仑在用扭秤测定两带电球间的作用力时,发现两带电球的作用力的定量关系与牛顿万有引力定律的数学关系相似,于是就把库仑力的定量关系类比于万有引力公式而得

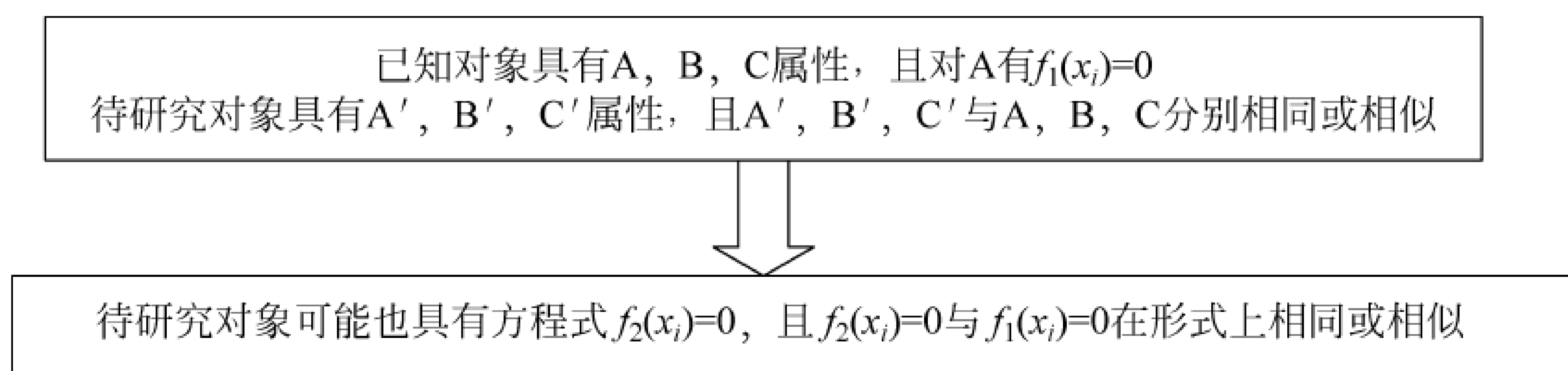



图 2-5-3　数学形式的类比一

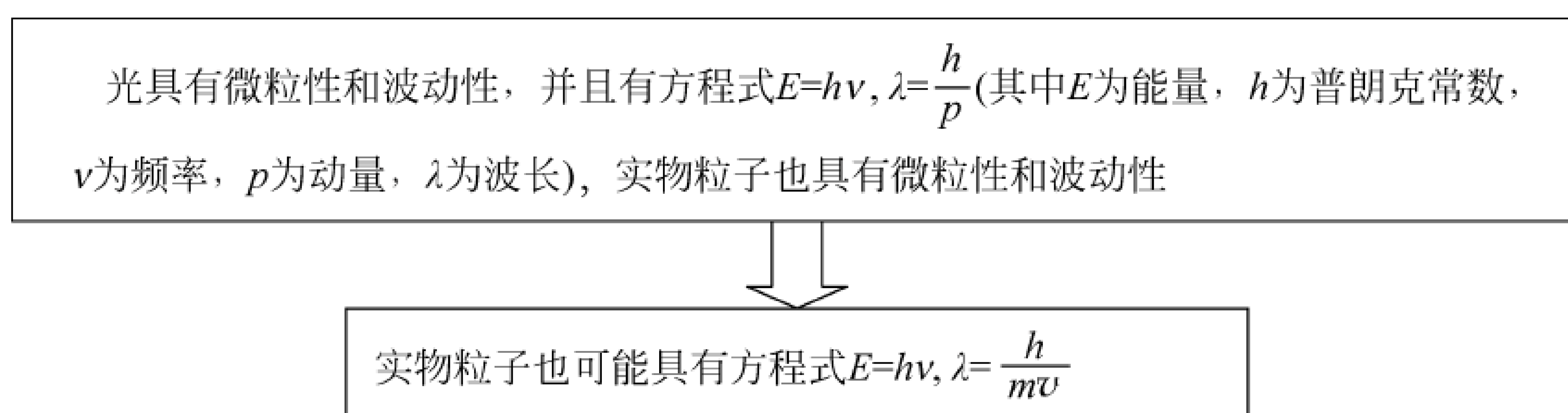



图 2-5-4　物质波公式的导出

出 $F=k\dfrac{q_1q_2}{r^2}$。

② 根据两个(或两类)对象有若干属性相同或相似，并且在两者数学方程式相同或相似的情况下，推论出它们其他方面的属性相同或相似。其基本模式如图 2-5-5 所示。

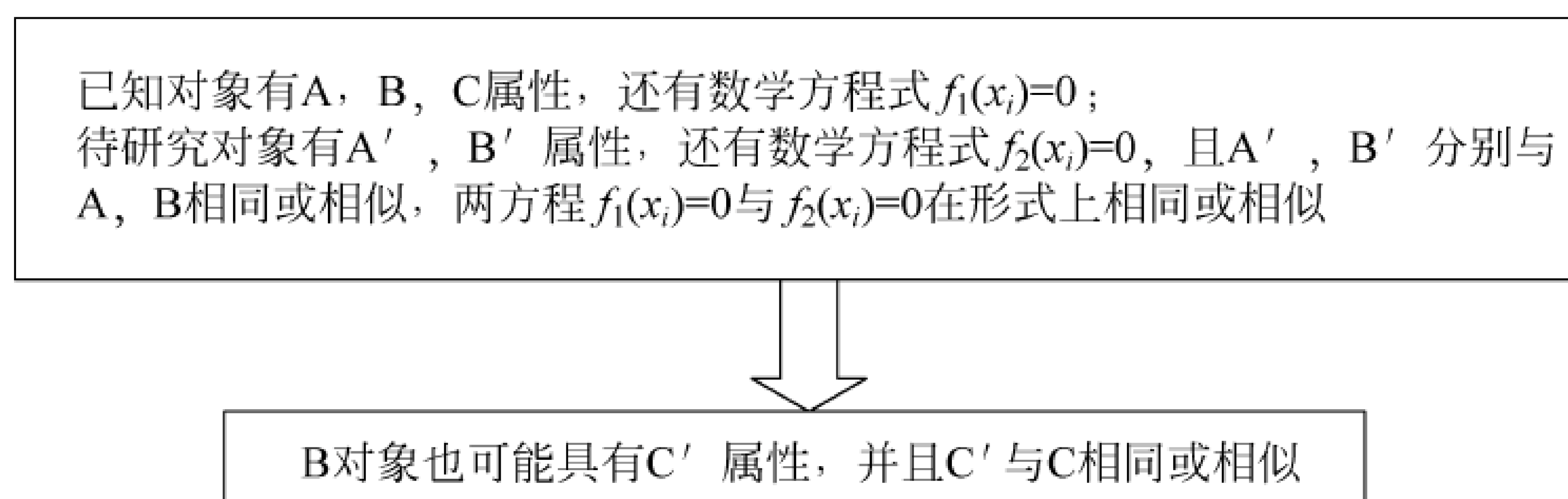



图 2-5-5　数学形式的类比二

例如，德布罗意运用数学形式类比推论出物质粒子也具有波动性的推理过程如图 2-5-6 所示。

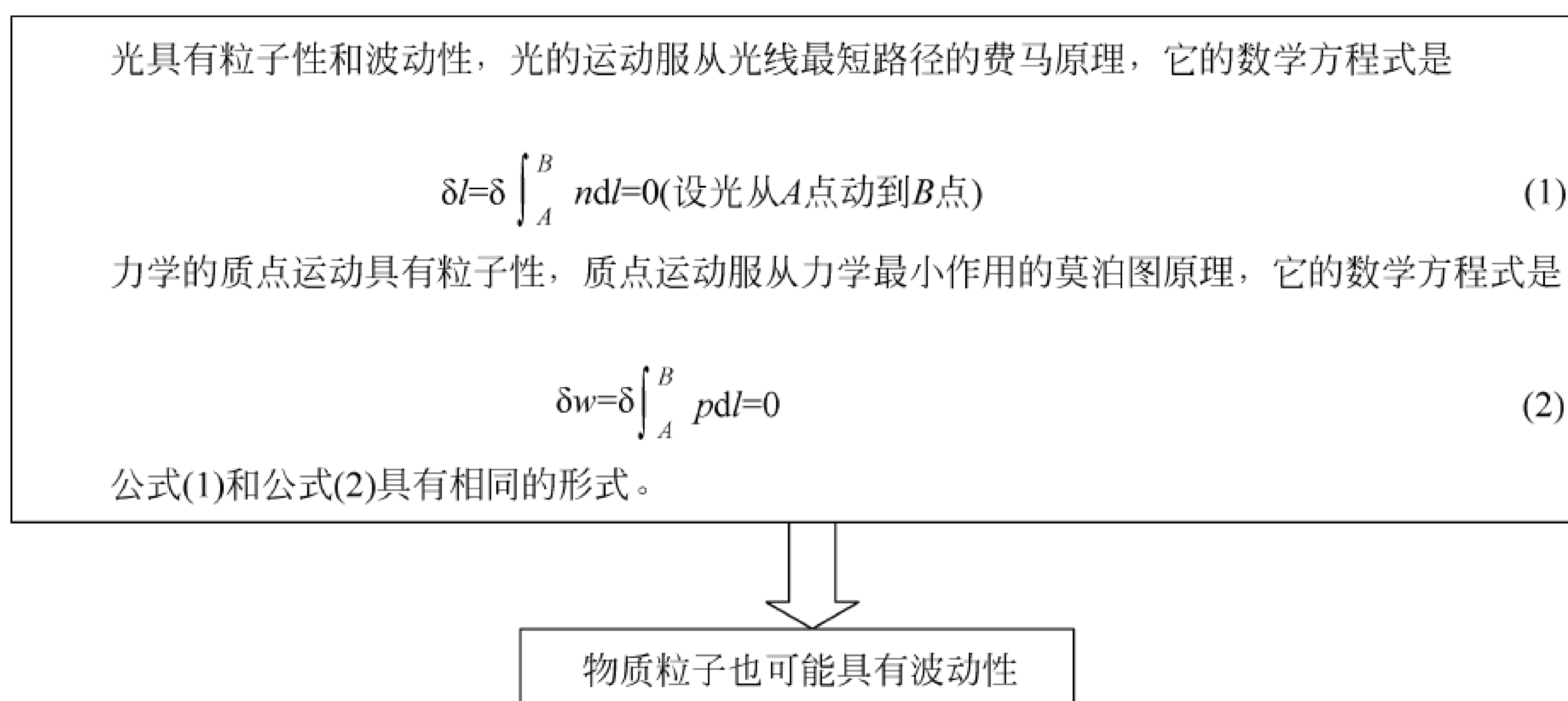



图 2-5-6　物质粒子具有波动性的推理过程

(3) 对称类比

这种形式的类比是根据两个(或两类)属性之间的对称关系进行的类比。客观世界中存在着许多对称性关系,这是进行对称类比的基础,对称类比的基本模式如图 2-5-7 所示。

例如,1931 年狄拉克(Paul Adrie Maurice Dirac,1902—1984)就是根据物质的对称关系,通过对称类比,提出了"正电子"的著名假说。在描述自由电子运动的狄拉克方程中,得出了两个正负对称的能量解。已知其中的正能量解对应负电子,那么"负能量解"应与谁对应呢?狄拉克将电子和电荷进行对比,既然有正负对称的两种电荷,也应该会存在与负电子对称的正电子。据此,狄拉克提出了正电子的存在问题,不久就被实验所证实。

(4) 简单共存类比

这种类比以简单共存关系作为中介。所谓简单共存关系,是指对象各个属性之间的关系仅在于它们都是同一对象的属性。它们之间可能是并列的、孤立的,也可能是人们尚不知道它们之间有无其他关系。这种类比的基本模式如图 2-5-8 所示。

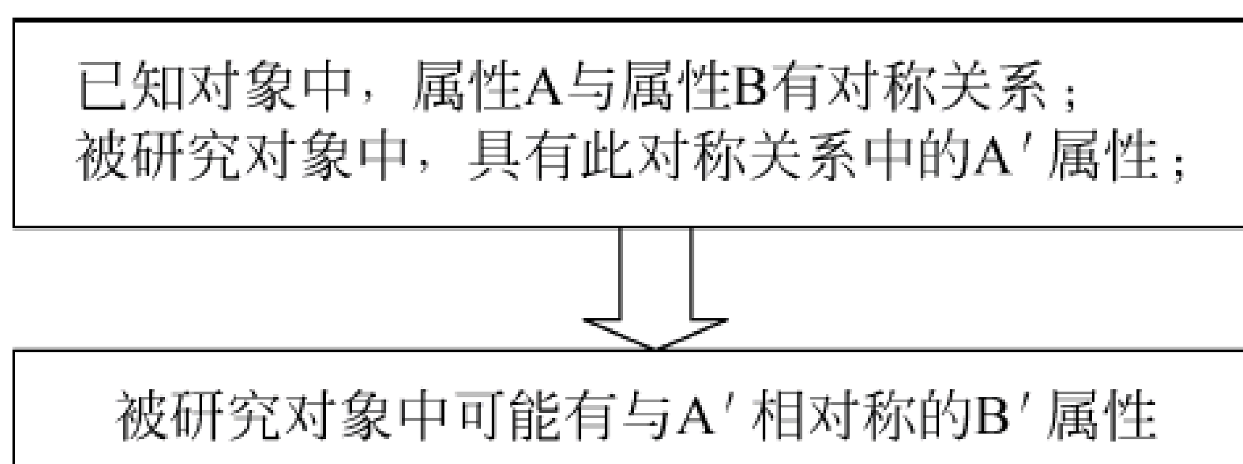



图 2-5-7 对称类比的基本模式

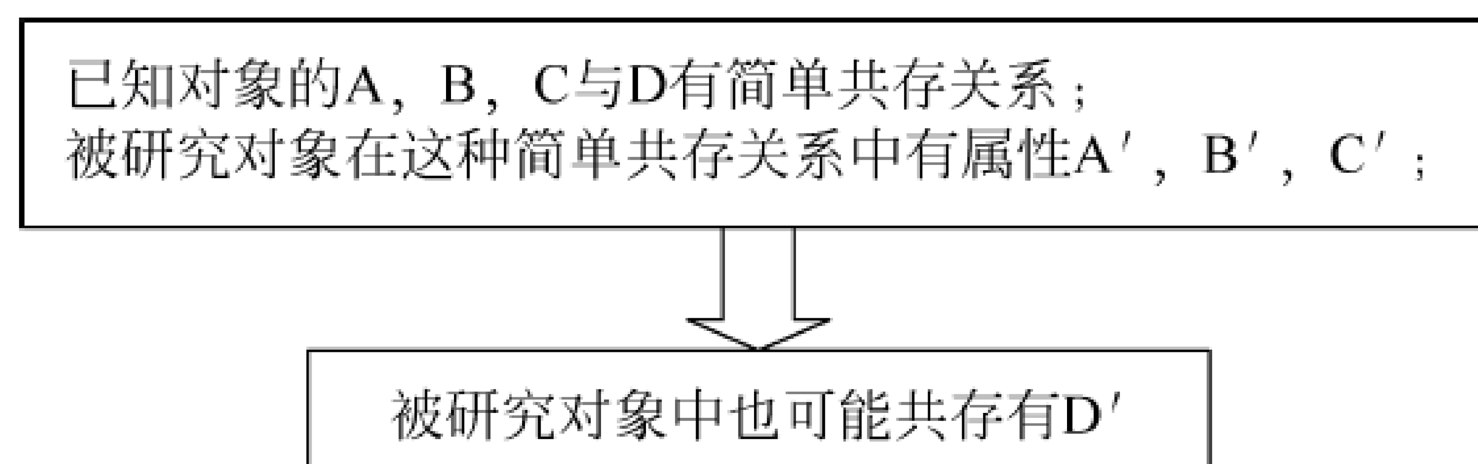



图 2-5-8 简单共存类比基本模式

这种形式的类比的应用比较多。例如在研究光现象的特性时,人们依据声现象与光现象特性的简单共存关系类比,从而得出"光也可能具有波动特性"的结论,并在后来被研究和实验所证实。再比如说,我们教材中还常常将电场和磁场进行简单共存的类比来推知磁场也是一种物质。

**3. 类比法的使用**

如前所述,类比法为我们探索物理规律提供了一种有利的工具。仅仅根据研究对象某些方面的相同或相似,就推论对象的其他方面也相同或相似,它推理的根据是不充分的,由此也必然导致其结论的或然性。因此在使用类比法时必须要注意以下几个方面:

(1) 要积累丰富的有关类比对象的材料

通常,类比的对象都是在研究者研究某一具体问题时,在思考的过程中自然浮现于脑际的。浮现的类比对象是否恰当、是否妥帖,与研究者知识积累密切相关。而且,确定类比对象之后,如果研究者知识广博、精深,就可以尽可能全面、透彻地把握研究对象的同一性和差异性,从而使类比的基础尽可能扎实,类比的结论尽可能符合客观实际。因此,在使用类比法之前必须对类比对象有全面深入的认识,这是使用类比法的一个必要条件。

(2) 在掌握类比对象的相似性和相关性的同时,还要注意类比对象的差异性

由于事物之间的相似性和相关性是类比的基础,所以,在使用类比法时,人们很自觉地会去对类比对象的相似性和相关性进行深入的研究,但类比对象的差异性往往被人们所忽视。当年惠更斯提出波动学说时,他通过发现"光"与"声"这两种不同的对象之间具有一系列的相似属性,即声音是借助于看不见摸不着的空气向声源周围的整个空间传播,最后到达我们的耳朵,光无疑也是从发光体(光源)通过某种传播媒介物质的运动而到达我们的眼睛,这就使得光与声具有类比的前提和基础,因而惠更斯类比到:光像声音一样也一定是以球

面波的形式来传播。然而，惠更斯在应用类比法时没有充分注意到“光”与“声”之间的不同之处和类比推理的结论具有或然性这一特点，从而得出了“光与声一样也是纵波”这一错误的类比结论。这个例子告诉我们：我们在不同的对象之间进行类比时，除了尽力找到类似的地方进行类比，提出新的假说或模型以外，还应当尽力地找到类比对象的不同点，以便适应新的假说或模型，做出创造性的发展。甚至可以说，由类比法所得出的结论，其可靠性应该取决于类比对象之间差异发现的程度，如果找不到或忽视了这种差异，由类比法得出的结论就会减低甚至完全失去其作用和价值。

### 2.5.2　模拟法

在物理实验中，若有两个物理实验的现象和过程具有相似性质，服从同一自然规律，满足同一边值条件，就可以利用其相似性研究其中一个现象以代替另一个现象，得到的结果是相同的，这种方法称为模拟法。

**1. 模拟法的分类**

根据模型和原型之间的相似关系，模拟法可分为物理模拟和数学模拟两种。物理模拟就是保持同一物理本质的模拟，例如用光测弹性模拟工件内部应力的分布等。

物理模拟可以分为三类(图 2-5-9)：几何模拟、动力相似模拟、替代或类比模拟(包括电路模拟)，这些方法在教学和工业中都有广泛的应用。数学模拟是把不同本质的物理现象或过程，用同一个数学方程来描绘。对于一个稳定的物理场，若它的微分方程和边界条件一旦确定，其解是唯一的。两个不同本质的物理场，如果描述它们的微分方程和边界条件相同，则它们的解是一一对应的，只要对其中一种易于测量的场进行测绘，并得到结果，那么与它对应的另一个物理场的结果也就知道了。

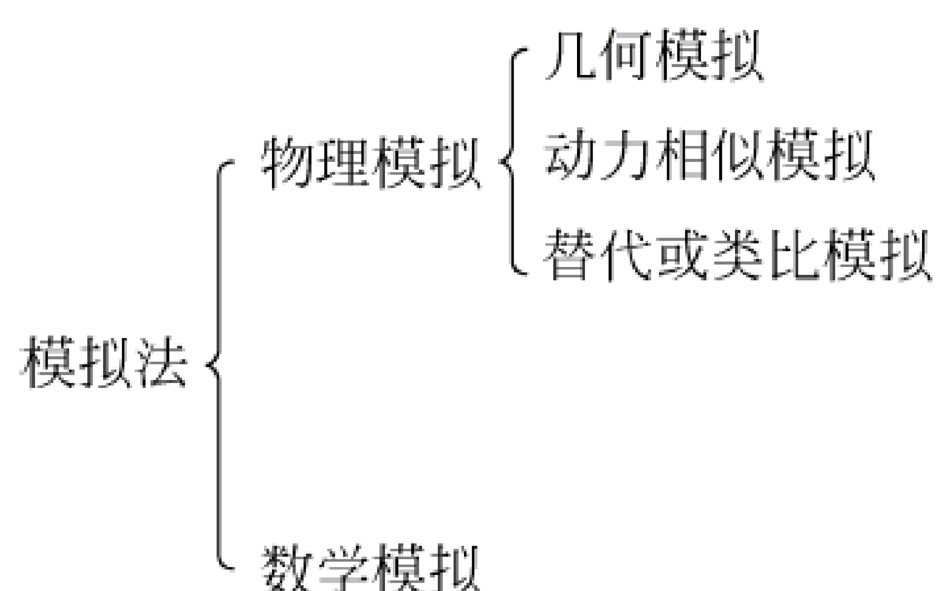



图 2-5-9　模拟法的分类

**2. 模拟法的作用**

模拟法本质上是用一种易于实现、便于测量的物理状态或过程来模拟不易实现、不便测量的物理状态或过程。在探求物质的运动规律和自然奥妙或解决工程技术或军事问题时，常常会遇到一些特殊的、难以对研究对象进行直接测量的情况。例如，被研究的对象非常庞大，比如：巨大的原子能反应堆、同步辐射加速器、航天飞机、宇宙飞船等；被研究的对象非常微小，比如：物质的微观结构、原子和分子的运动等；还有被研究的对象非常危险，比如：地震、火山爆发、发射原子弹或氢弹等；或者是研究对象变化非常缓慢，比如：天体的演变、地球的进化等。根据相似性原理，可人为地制造一个类似于被研究的对象或者运动过程的模型来进行实验。因此，模拟法是一种重要的实验研究方法。

**3. 模拟法的应用**

下面用稳恒电流场模拟静电场为例说明模拟法的应用。

1）用稳恒电流场模拟静电场的理论依据

静电场是由电荷分布决定的，确定静电场的分布，对于研究带电粒子与带电体之间的相互作用是非常重要的。理论上讲，如果知道了电荷的分布，就可以确定静电场的分布。在给定条件下，确定系统静电场分布的方法，一般有解析法、数值计算法和实验法。在科学研究和生产实践中，随着静电应用、静电防护和静电现象等研究的深入，常常需要了解一些形状比较复杂的带电体或电极周围静电场的分布，这时，用理论方法(解析法和数值计算法)是十分困难的。

然而,对于静电场来说,要直接进行探测也是比较困难的。其一是,静电场中无电流,一般的磁电式仪表不起作用,只能用静电式仪表进行测量,而静电式仪表不仅结构复杂,而且灵敏度也较低;其二是,仪表本身是由导体或电介质制成的,静电探测的电极一般很大,一旦放入静电场中,将会引起原静电场的畸变。

电介质中的稳恒电流场与静电场是两种不同性质的物理场,但是它们在一定条件下具有相似的空间分布,即两种场有相似的物理规律。另外,由于标量的测量和计算比矢量简便,因此,人们更愿意用电势来描述电场。

静电场与稳恒电流场的对应关系如表 2-5-1 所示。

**表 2-5-1 静电场与稳恒电流场的对应关系**

| 静　电　场 | 稳恒电流场 |
|---|---|
| 导体上的电荷$\pm Q$ | 极间电流 $I$ |
| 电场强度 $\boldsymbol{E}$ | 电场强度 $\boldsymbol{E}$ |
| 介电常量 $\varepsilon$ | 电导率 $\sigma$ |
| 电位移 $\boldsymbol{D}=\varepsilon\boldsymbol{E}$ | 电流密度 $\boldsymbol{J}=\sigma\boldsymbol{E}$ |
| 无荷区 $\oint\varepsilon\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{S}=0$ | 无源区 $\oint\sigma\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{S}=0$ |
| 电势分布$\nabla^2 U=0$ | 电势分布$\nabla^2 U=0$ |

由此可见,这两种物理场有一一对应的物理量,并且满足相同形式的数学规律,这是可以用稳恒电流场模拟静电场的前提条件。另外,对于稳恒电流场,只有改变其电压的大小或把介质置入电流场中,从而改变电流大小或流向,才会引起电流场中电势的改变。也就是说,稳恒电流场比静电场更稳定、更易测量,这就是用稳恒电流场模拟静电场的主要原因。

2) 同轴电缆的静电场分布及同轴圆柱面电极间的电流分布

为了分析问题的需要,现将同轴电缆的静电场分布及同轴圆柱面电极间的电流分布进行理论上的比较。如图 2-5-10(a) 所示,在真空中有一半径为 $r_a$ 的长圆柱体 A 和一内半径为 $r_b$ 的长圆筒形导体 B,它们同轴放置,分别带等量异号电荷。由高斯定理知,在垂直于轴线的任一截面 $S$ 内,都有均匀分布的辐射状电场线,这是一个与坐标 $z$ 无关的二维场。在二维场中,电场强度 $E$ 平行于 $xOy$ 平面,其等位面为一簇同轴圆柱面。因此,只要研究 $S$ 面上的电场分布即可。

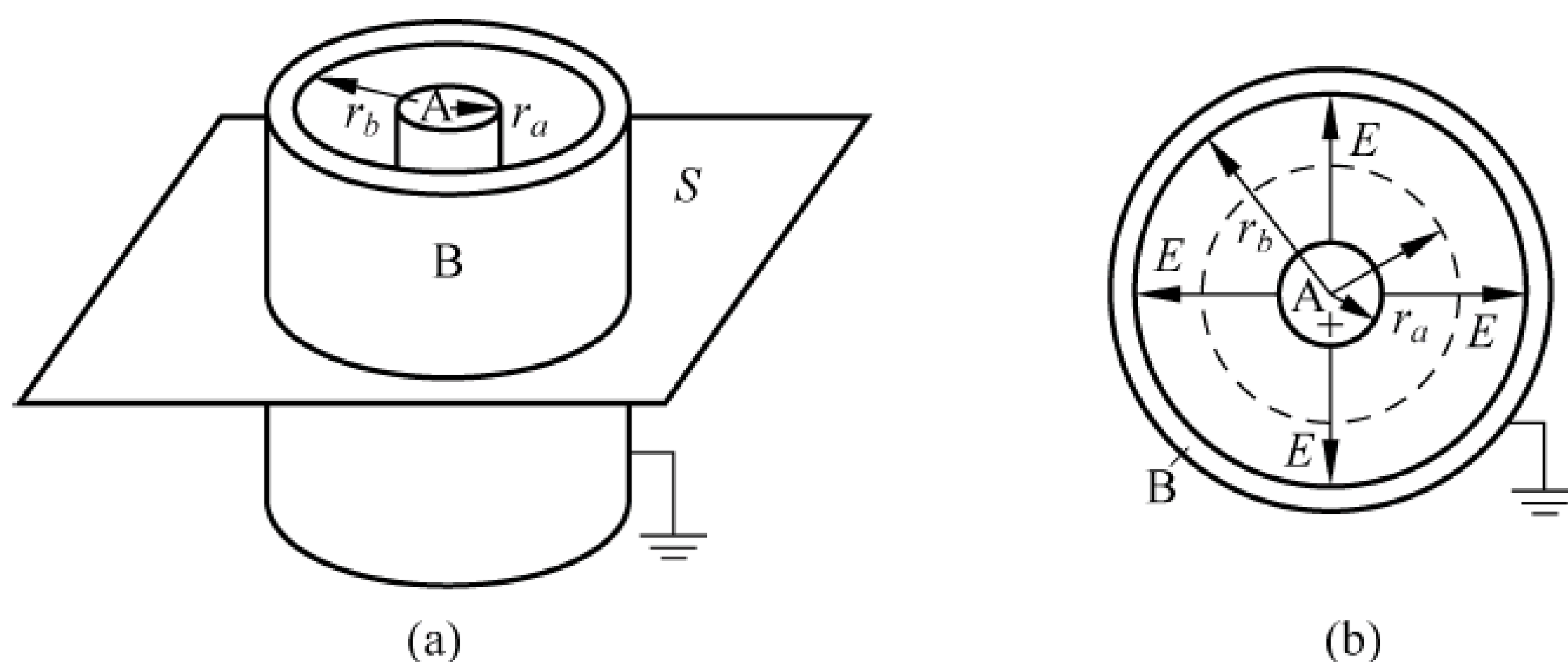



图 2-5-10 同轴电缆及其静电场分布

由静电场中的高斯定理可知,距轴线的距离为 $r$ 处,图 2-5-10(b)的各点电场强度为 $E=\dfrac{\lambda}{2\pi\varepsilon_0 r}$,式中 $\lambda$ 为柱面每单位长度的电荷量,其电位为

$$U_r = U_a - \int_{r_a}^{r} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{r} = U_a - \frac{\lambda}{2\pi\varepsilon_0}\ln\frac{r}{r_a}$$

设 $r=r_b$ 时，$U_b=0$，则有$\frac{\lambda}{2\pi\varepsilon_0}=\frac{U_a}{\ln r_b/r_a}$，得

$$U_r = U_a \frac{\ln r_b/r}{\ln r_b/r_a} \tag{2-5-1}$$

和

$$E_r = -\frac{\mathrm{d}U_r}{\mathrm{d}r} = \frac{U_a}{\ln r_b/r_a} \cdot \frac{1}{r} \tag{2-5-2}$$

若上述圆柱形导体A与圆筒形导体B之间充满了电导率为$\sigma$的不良导体，A，B与电源电流正负极相连接(见图2-5-11)，A，B间将形成径向电流，建立稳恒电流场$E_r$，可以证明不良导体中的电场强度$E'_r$与原真空中的静电场$E_r$是相等的。

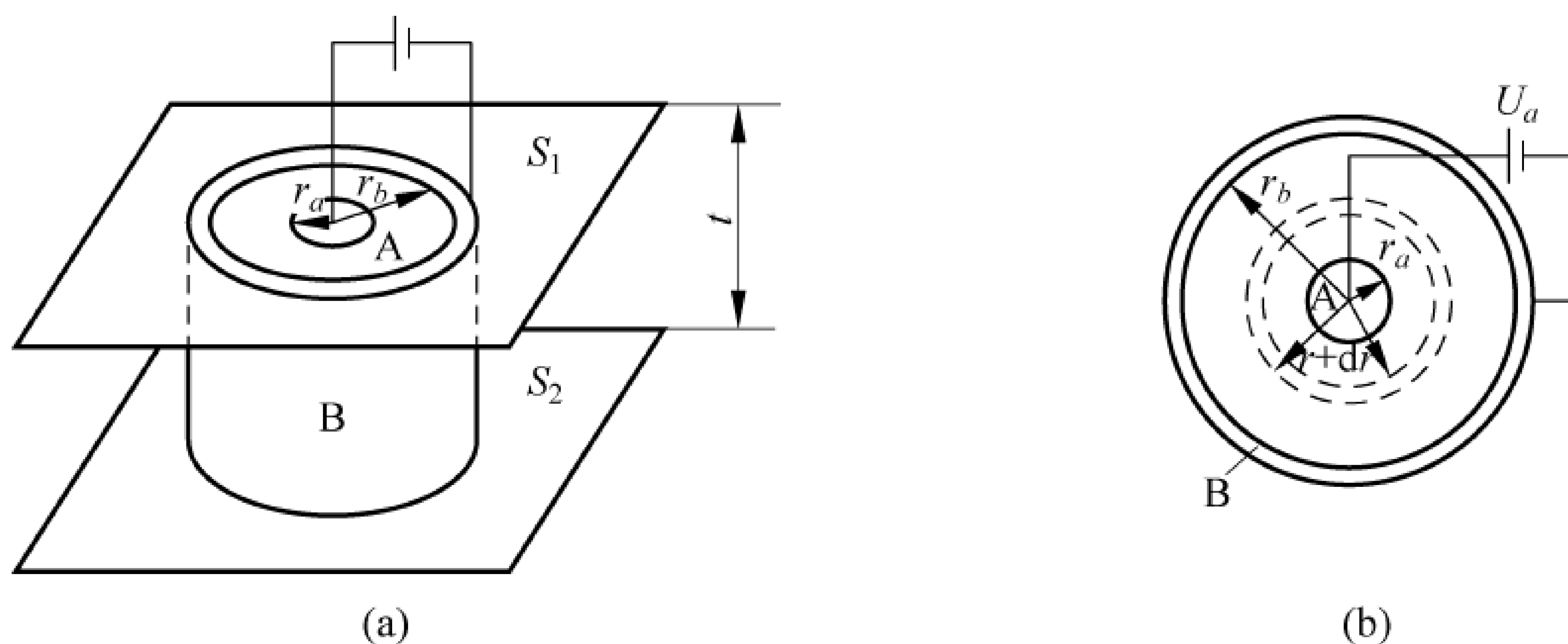



图2-5-11 同轴电缆的模拟模型

取厚度为$t$的圆轴形同轴不良导体片为研究对象，设材料电阻率为$\rho\left(\rho=\frac{1}{\sigma}\right)$，则任意半径$r$到$r+\mathrm{d}r$的圆周间的电阻是$\mathrm{d}R=\rho\cdot\frac{\mathrm{d}r}{s}=\rho\cdot\frac{\mathrm{d}r}{2\pi rt}=\frac{\rho}{2\pi t}\cdot\frac{\mathrm{d}r}{r}$，则半径为$r$到$r_b$之间的圆柱片的电阻为$R_{rr_b}=\frac{\rho}{2\pi t}\int_{r}^{r_b}\frac{\mathrm{d}r}{r}=\frac{\rho}{2\pi t}\ln\frac{r_b}{r}$。总电阻为(半径$r_a$到$r_b$之间圆柱片的电阻) $R_{r_ar_b}=\frac{\rho}{2\pi t}\ln\frac{r_b}{r_a}$。设$U_b=0$，则两圆柱面间所加电压为$U_a$，径向电流为$I=\frac{U_a}{R_{r_ar_b}}=\frac{2\pi tU_a}{\rho\ln r_b/r_a}$，则距轴线$r$处的电位为$U'_r=IR_{rr_b}=U_a\frac{\ln r_b/r}{\ln r_b/r_a}$。则

$$E'_r = -\frac{\mathrm{d}U'_r}{\mathrm{d}r} = \frac{U_a}{\ln r_b/r_a} \cdot \frac{1}{r} \tag{2-5-3}$$

由以上分析可见，$U_r$与$U'_r$，$E_r$与$E'_r$的分布函数完全相同。两种不同性质的物理场却具有相同的分布，这可以从电荷产生场的观点加以分析。在导电质中没有电流通过时，其中任一体积元(宏观小，微观大，其内仍包含大量原子)内正负电荷数量相等，没有净电荷，呈电中性。当有电流通过时，单位时间内流入和流出该体积元内的正或负电荷数量相等，净电荷为零，仍然呈电中性。因而，整个导电质内有电流通过时也不存在静电荷。这就是说，真空中的静电场和有稳恒电流通过时导电质中的场都是由电极上的电荷产生的。事实上，真空中电极上的电荷是不动的，在有电流通过的导电质中，电极上的电荷一边流失，一边由电源

补充，在动态平衡下保持电荷的数量不变。所以这两种情况下电场分布是相同的。

3）模拟静电场的实验方法

由以上的理论分析，可以看出稳恒电流场和静电场有一一对应的物理量，但要想在实验上用稳恒电流场来模拟静电场，需要满足下面三个条件：

（1）稳恒电流场中的电极形状应与被模拟的静电场中的带电体几何形状相同；

（2）稳恒电流场中的导电介质是不良导体且电导率分布均匀，并满足 $\sigma_{电源} \gg \sigma_{导电源}$ 才能保证电流场中的电极（良导体）的表面也近似是一个等位面；

（3）模拟所用电极系统与模拟电极系统的边界条件相同。

场强在数值上等于电位梯度，方向指向电位降落的方向。考虑到场强是矢量，而电位 $U$ 是标量，从实验测量来讲，测定电位比测定场强容易实现，所以可先测绘等位线，然后根据电场线与等位线正交的原理，画出电场线。这样就可由等位线的间距确定电场线的疏密和指向，将抽象的电场形象反映出来。

为了达到模拟的条件，实验中采用了 GVZ-3 型导电微晶静电场描绘仪（包括导电微晶、双层固定架和同步探针等），实验装置如图 2-5-12 所示。支架采用双层式结构，上层放记录纸，下层放导电微晶。电极已直接制作在导电微晶上，并将电极引线接出到外接线柱上，电极间的导电率远小于电极且各项均匀的导电介质。接通直流电源（10V）就可以进行实验。在导电微晶和记录纸上方各有一探针，通过金属探针臂把两探针固定在同一手柄座上，两探针始终保持在同一铅垂线上。移动手柄座时，可保证两探针的运动轨迹是一样的。由导电微晶上方的探针找到待测点后，按一下记录纸上方的探针，在记录纸上留下一个对应的标记。移动同步探针在导电微晶上找出若干电位相同的点，由此即可描绘出等位线。

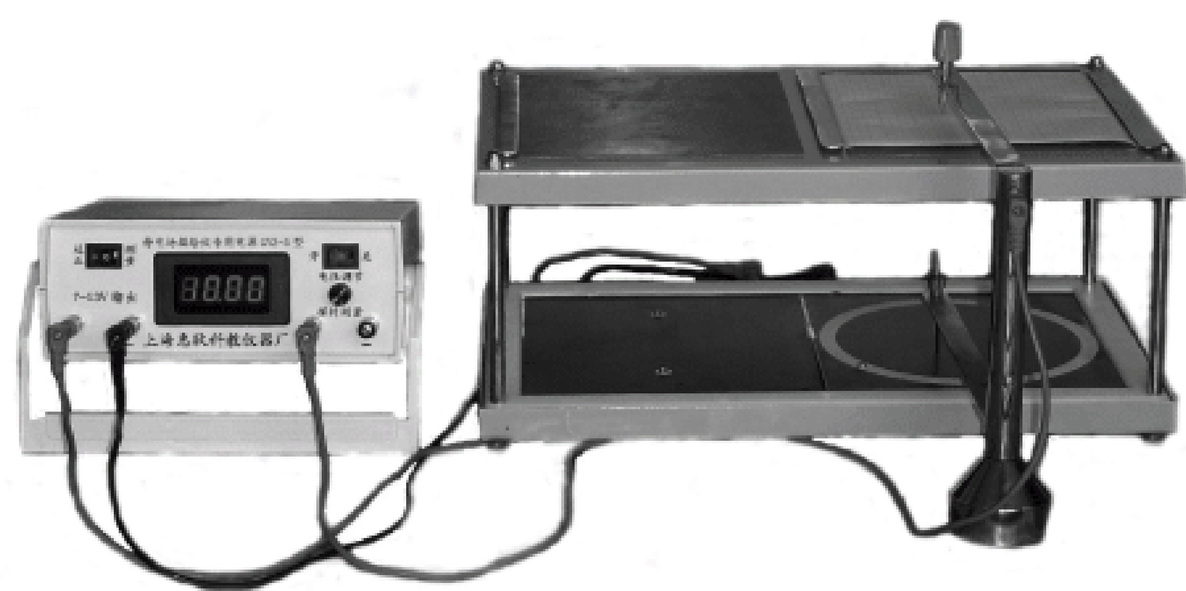


图 2-5-12 GVZ-3 型导电微晶静电场描绘仪

4）模拟法的优缺点

模拟法作为科学研究的一种基础的方法，其应用范围越来越广。模拟法的主要优点有：可以对已经时过境迁或尚未出现的现象进行研究；可以对那些既不能打开，又不能从外部直接观察其内部状态的系统，进行研究；可将现象简化、放大或缩小；易于控制；比较经济。但是，它也有缺点，主要是人工模仿和复制的人为性，难免使得出的结论欠准确，欠完整，不一定符合所模拟的对象。特别值得注意的是，模拟法的使用有一定的条件和范围，不能随意推广，否则会得到荒谬的结论。上述的例子之所以可以用稳恒电流场来模拟静电场，就是因为这些物理场具备了可模拟的条件，即有一一对应的物理量，且满足相似的数学形式及边界条件。

## 2.6 归纳与演绎

众所周知，牛顿在伽利略、开普勒等前人成果的基础上，用归纳法获得了经典力学的基本概念和力学三大定律，又用演绎的方法获得万有引力定律并发明了微积分。在物理学研究的常用方法中，经常用到归纳与演绎的思维方法。应当说，归纳和演绎是人类认识最早、运用最为广泛的思维方法。它所涉及的是个别与一般的关系，是事物和概念之间的外部关系。归纳与演绎这两种从某种意义上是互相独立的思维方法，实际上是辩证统一的。

## 2.6.1　归纳和演绎的含义

**1. 归纳的含义**

所谓归纳，是指从许多个别的事物中概括出一般性概念、原则或结论的思维方法。

归纳可分为完全归纳法和不完全归纳法。完全归纳法是前提包含该类对象的全体，从而对该类对象做出一般性结论的方法。不完全归纳法又称简单枚举归纳法，是通过观察和研究，发现某类事物中固有的某种属性，并且不断重复而没遇到相反的事例，从而判断出所有该类对象都有这一属性的推理方法，数学上的穷举法就是完全归纳法。简单枚举归纳法的结论带有或然性，可能为真，也可能为假。

在实践中，人们首先获得一些个别事物的知识，然后在这些特殊性知识的基础上，概括出同类事物的普遍性知识。比如，人们从宏观世界万物都可分为若干层次，微观世界的原子可再分为基本粒子以至夸克等事实，得出“物质是无限可分的”一般原理。这个认识过程就包含着归纳推理。

从个别事实出发，推出普遍性结论的方法。比如，为了研究宇宙飞行时的超重和失重（主要是失重）状态对动物身体的影响，科学家们先用个别的动物（如狗、老鼠和果蝇）进行实验，观察它们在失重状态下的反应以及经过宇宙飞行后身体的变化。经过许多次实验以后，科学家就从这些个别的动物能够经受长时间的失重状态推理出一条普遍性的规律——“所有的动物都能经受长时间的失重状态”。以后，科学家们又让个别的宇航员进行相应的实验，发现这些宇航员也能经受长时间的失重状态。于是，科学家们又得出一条普遍性的规律——“人类能够适应较长时间的失重状态”。

从个别事物的性质中概括出共同本质的方法也是归纳法。人们对电现象的认识也经历了这样的历程。起初，人们接触到各种各样的、特殊的电现象：夏日的天空，乌云翻滚，电闪雷鸣，人们称之为雷电；非洲和南美的土人对一种现在称为电鳗的热带鱼很害怕，渔民在捕捉它们的时候，鱼会突然放电，以一种莫名其妙的可怕力量，狠狠地打击捕捉它的人，这种现象人们称之为生物电；摩擦橡胶棒，会使棒具有吸引轻小物体的性质，人们称之为摩擦电；伏打电池产生的电，称为化学电；电磁感应现象产生的电，称为磁电……以后，人们对这些电现象进行了比较和概括，发现它们都有共同的本质——都是由于电子的得失造成的电现象。

**2. 演绎的含义**

演绎，是以一般概念、原则为前提推导出个别结论的思维方法，即依据某类事物都具有的一般属性、关系来推断该类事物中个别事物所具有的属性、关系的推理方法。

如果掌握了某一类事物的共同属性或某种必然性的知识，就可以将它推广到这类事物的任何个别事物，从而引出个别结论，这就是演绎方法。比如，人们根据“物质是无限可分的”原理，推知基本粒子也是可分的，这就是演绎推理得来的。又如，人皆有死，苏格拉底是人，所以，苏格拉底也是有死的。再如：水果都含维生素，梨是水果，所以梨含维生素。演绎推理是一种必然推理，凡大前提正确，小前提无误，推理符合逻辑，结论一般正确。

伽利略用演绎法推翻亚里士多德的“物体越重，下落速度越快”的论断。

## 2.6.2　归纳与演绎的相互关系

**1. 归纳与演绎的相互独立性**

在近代科学形成的同时，就有哲学家思考：科学是什么？科学的方法是什么？像英国

的哲学家培根,他是经验论者,强调归纳法,他认为科学定律是通过归纳得出的,他提出了一个著名口号:“知识就是力量”,强调:搞科学就是要征服自然,控制自然。另外法国的笛卡儿,他是理性论者,他写过一本书叫作《方法谈》,书中他强调理性的逻辑的演绎方法。这是两种科学观,正如爱因斯坦所讲的,培根强调的是外部证实,笛卡儿强调的是内部的逻辑的完备。

归纳与演绎是可以相互独立的。

在20世纪科学革命以后,实证论有很大的发展,称为逻辑经验论,或者叫逻辑实证论。这就是维尔纳学派的哲学。维尔纳学派的领袖是史里克,他原来是学物理的,博士论文是关于相对论的,但是他后来搞科学哲学。那个时候他们对归纳法已经有些看法,就是归纳法并不能归纳出一个普遍规律来。实际上要从经验材料中得出一个规律、一个定理,并不是靠归纳法就能归纳得出来的。比较简单的定理也许可以归纳出来,但像相对论、量子力学这样的理论绝对不是归纳法能归纳出来的。所以,后来物理学家用的是一种假说演绎法,根据一些经验事实,通过自由地创造提出一个假说,然后由假说再通过演绎推理,推理出一些可以用实验和观测来检验的推论。这些推论假如证实了,那么说明这个理论是对的;推论证实不了,那么这个理论就不对,要重新修正。这就是假说演绎法。假说演绎法里面很根本的问题就是证实,所以,逻辑经验论者就强调,我们是科学的哲学,我们不要那种没有办法证实的形而上学。所以他们要取消形而上学,强调一切东西都要拿来检验,要证实。一切东西都要拿出证据来,这是他们的一个核心的思想,也是科学与非科学、伪科学划界的一个标准。就是说,你的理论是可以用实验用观察证实的就是科学,假如不能证实的就不是科学。但是后来证实原则在他们这些逻辑经验论者里边进行了很深入的反复争论,发现许多理论和定律没法证实。特别是一些全称命题、普遍性命题是没法证实的。例如,天下的乌鸦都是黑的,这是个普遍性的判断。但是这个普遍性的判断,用归纳法讲,今天我们抓到的乌鸦是黑的,明天抓到的乌鸦是黑的,那是不是证实了呢?还是没有证实。因为你没有把天下所有的乌鸦都抓起来,所以你还是不知道天下的乌鸦是不是都是黑的。所以说普遍命题是没法证实的。关于这个问题,后来做了很多讨论,很多修正,这个证实原则争论了很长时间,这个证实原则解决不了。只能够说我们做实验支持这个结论,支持这个假说,支持这个理论,确认这个理论,确认度高一点或低一点。但是我们没有办法绝对地说,这个定理就是绝对百分之百正确的,因为你不可能把所有东西都检查过。

另一方面,我们又要注意,如果说科学需要创新,强调演绎也是必要的。例如,我们看看爱因斯坦的科研经历。从爱因斯坦的科研经历可以看出,他特别看重逻辑的演绎法。这对那些认为逻辑思维限制了创新的观点来说,无疑是一个打击。拉普拉斯说:“认识一位巨人的研究方法,对于科学的进步来说是至关重要的。”人们要研究某个问题,总得先收集有关资料,尽量汲取前人的成果、方法和经验,这样才能站在“巨人的肩上”(牛顿语)。但为了比前人看得更远,光站在巨人的肩上还不够,还需要通过新的观察或试验,向大自然索取更多的新资料,并运用辩证思维,对它们进行分析整理,以获得新的认识。至此,爱因斯坦与常人并无显著不同,他和我们一样,也非常重视基本理论的学习,重视来自实际的实验和事实。然而,在如何对资料进行理论分析时,则与常人有所不同。常见的办法是把试验结果归纳成一些经验定律或经验公式,希望通过它们合理地解释过去并预见将来。但爱因斯坦认为这样做还很不够,这样做不能使理论获得重大进展。因为,这些公式最多只概括了局部的、多次有限的经验,对全局来说,对无限的总体来说,它很可能是错误的。从学习方法上看,爱因斯

坦也有出众之处。他善于“识别出那种能导致深邃知识的东西，而把其他许多东西撇开不管；在他所阅读的书本中找出可能把自己引到深处的东西，把其他一切统统抛掉，就是使头脑负担过重和会把自己远离要点的一切都抛掉。”这样，他就把全部精力集中在最能发挥自己的创造性与能动性的问题上了。爱因斯坦的这种科学研究方法为当今的科技创新提供了宝贵的借鉴。因为它忽略了直觉思维、演绎思维等逻辑思维方法在精密科学发展中所起的重大作用。爱因斯坦说：近代自然科学时期是“以归纳为主的方法”，应让位于“探索性的演绎法”；应该“由经验材料作为引导提出一种思想体系，它一般是在逻辑上从少数几个所谓公理的基本假定建立起来的”。对这个体系的要求，应是能把观察到的事实联结在一起，同时它还具有最大可能的简单性：所谓简单性是指“这体系所包含的彼此独立的假设或公理最少”。在爱因斯坦的科研成果中，其定理或公理的数量并不多，就狭义相对论而言其公理只有两条，即相对性原理和光速不变原理。至于思想体系的内容，它应由“概念、被认为对这些概念是有效的基本定律，以及用逻辑推理得到的结论这三者所构造的”，也就是通常所说的概念、公理和定理三部分。几何学与代数是最具演绎性的科学体系。作为自然科学史上的第一个思想体系的光辉例子就是欧几里得几何学。如果欧几里得等人当年只满足于把丈量土地所得的具体结果，归纳为若干条经验定律，那么，几何学的发展也许会延误许多。等等这些，都可以看到关注演绎的重要性。

**2. 归纳与演绎的辩证统一**

归纳与演绎方法不是对立的，而是辩证统一的，对于这个观点我们可以从牛顿的成功原因分析上看到：牛顿在物理学方面取得非凡的成就源于他对自然科学的浓厚兴趣及实验运用归纳法与演绎法相结合的研究方法。

归纳和演绎反映了人们认识事物两条方向相反的思维途径，前者是从个别到一般的思维运动，后者是从一般到个别的思维运动。

归纳和演绎是形式逻辑和辩证逻辑共有的思维方法，是辩证思维的起点。所不同的是，形式逻辑把归纳和演绎看作是各自独立、相互平行的两种逻辑的证明工具和推理规则，割裂了归纳和演绎的辩证关系，并且，形式逻辑抛开事物的具体内容和矛盾，只注重归纳和演绎的形式，因而总是从不变的前提出发，按照固定的线路，推出僵硬的结论。与形式逻辑相反，辩证逻辑强调归纳和演绎是既相互区别，又相互联系的两种思维方法，是概念、理论形成过程不可分割的两个侧面。

首先，归纳与演绎相互联系、互为条件。一方面，没有归纳就没有演绎，归纳是演绎的基础，为演绎提供前提。演绎要从一般推导出个别，作为演绎出发点的一般原则，往往是先由归纳得出来的。例如，生物遗传的基因学说就是归纳了大量生物实验事实得出来的。又如，“人皆有死”作为演绎推理的前提，是从社会实践中归纳得出的结论。另一方面，没有演绎也没有归纳，演绎为归纳提供指导。归纳要从个别概括出一般，作为对实际材料进行归纳的指导思想，往往又是某种演绎的结果。例如，达尔文（Charles Robert Darwin，1809—1882）把大量观察、实验材料进行归纳，得出“生物进化”这个结论；但他在得出“生物进化”这个结论之前，早就接受了法国博物学家和生物学家拉马克（Jean Baptiste Lemarck，1744—1829）等人的有关生物进化的思想和赖尔（Martin Ryle，1918—1984）的地质演化思想，这些思想实际上构成了他归纳经验材料的指导原则，因为有了这些思想，达尔文的考察、归纳才显得有目的性和选择性。

其次，归纳和演绎相互补充、相互转化。这是由于在思维运动中，二者虽然都有重要作

用,但各自也都存在一定的局限性:归纳法只是对现存的有限的经验材料进行概括,因而不仅不能保证归纳结论的普适性,而且难以区分事物的本质属性与非本质属性,这就使得归纳推理的结论可能为真,也可能为假。演绎法从一般原则出发思考问题,但它无法保证自己的前提,即出发的一般原则本身是否正确无误。因此,归纳与演绎必须在相互转化过程中,弥补各自的缺陷。归纳之后,需要通过演绎将归纳所得的一般结论推广到未知的事实上,并用这些事实来检验一般结论的正确与否;演绎之后,又要将演绎所得的个别结论与事实相比较,并通过新的归纳来检验、修正、充实原有的演绎前提。归纳和演绎只有在如此周而复始的相互转化过程中,才能弥补各自的缺陷,充分发挥其在探索真理过程中的方法论作用。

### 2.6.3 归纳方法和演绎方法在物理学研究中的应用

**1. 实验运用归纳**

大约在1663年,牛顿即开始热衷于光学研究,磨玻璃、制作望远镜也在这个时期。1666年,牛顿购得一块玻璃三棱镜,开始研究色散现象。牛顿在他的《光学》一书写道:“把我的房间弄暗,在窗板上开个小孔,以便适量的太阳光射入室内,就在入口处安置我的棱镜,光通过棱镜折射达到对面的墙上。”牛顿看到墙上有色彩的光带,光带之长数倍于原来的白光点,他意识到这些彩色就是组成白色太阳光的原始色彩。为了证明这一点,牛顿进一步在光带投射的屏上也打一小孔。让光带中彩色的一部分穿过第二个小孔,经过放在屏后的第二个棱镜折射投到第二个屏上,又让第一个棱镜绕它的轴缓慢转动。只见穿出第二个小孔落在第二屏上的像随着第一棱镜转动而上下移动。于是看到,第一棱镜折射最大的蓝光,经过第二棱镜也是折射最大。反之,红光被前后两个棱镜折射最小。于是牛顿得出结论:经过第一棱镜折射所得的长方形的彩色光带不是别的,正是由不同的各种彩色光所组成的非均匀的混合体。这就是牛顿的光学色散理论,它是通过实验建立起来的。牛顿自称这个实验为“关键性实验”。这个实验可以说是一个半世纪后夫琅禾费建立光谱术的基础。

牛顿在《光学》一书中凭借实验的结果与分析,建立了光的理论。爱因斯坦在为牛顿《光学》1931年重印本所作的序中说:“牛顿的时代已被淡忘了……牛顿的各种发现已进入公认的知识宝库,尽管如此,它的《光学》著作的这个新版本还是应当受到我们怀着衷心感激的心情去欢迎的。因为只有这本书才能使我们有幸看到这位伟大人物本人的活动。”

**2. 逻辑演绎**

1713年,牛顿在出版《自然哲学的数学原理》第二版时在给他的学生科茨的信中提到,运动定律是居于首位的定律或称之为公理,说它们都是从现象中推断或称演绎而来的,并运用归纳法使之普适化,牛顿说:“这是一个命题在哲学中所能达到的最高境界的例证,必须看到归纳与演绎不能人为地对立起来。”

关于实验与假设之间的关系,牛顿在各种场合都有论述。他在给奥尔登堡(Oldenburg)的信中说:“进行哲学研究的最好和最可靠的方法,看来第一是勤勤恳恳地探索事物的属性并用实验来证明这些属性;然后进而建立一些假说用来解释这些事物的本性,牛顿这些论述奠定了自然哲学的基础,启开了实验科学的大门,300年来为自然科学的繁荣立下了不朽功勋。”

牛顿研究事物规律的方法不同于那些只从简单的物理假设出发的人,而是通过逻辑的演绎法得到对事物现象的解释。爱因斯坦指出:“牛顿第一个成功地找到了一个用公式清楚表达的基础,从这基础出发,他用数学的思维,逻辑、定量地演绎出范围很广的现象,并且

同经验相符合。”牛顿是完整的物理因果关系创造人，而因果关系正是经典物理学的基石。

运用归纳法和演绎法研究问题是走向成功的最普遍，也是最有效的方法。它不仅仅适用于科学研究，它还同样适用于整个人类社会的各种研究。

比如：已知均匀磁场（磁感应强度为 $B$）对长度为 $l$ 的载流导线（电流强度为 $I$）施以力 $F$ 的作用，其大小为 $F=IlB$。该式适用于载流导线垂直于磁场方向放置的情况。根据这一规律，可用演绎推理方法得出运动电荷在磁场中受力的大小。因为电流是自由电荷的定向运动，磁场对载流导线的作用力可看作是磁场对导线中定向运动的自由电荷的作用力总和。设每米导线中有 $n$ 个自由电荷，长度为 $l$ 的导线中共有 $nl$ 个自由电荷，每个电荷的电量为 $q$，电荷定向运动的速度为 $v$，根据电流强度的定义，可得：$I=nqv$。则 $F=IlB=nqvlB$，因此，磁场对每个垂直于磁场方向运动的自由电荷的作用力的大小为 $f=qvB$。

显然，演绎推理的出发点是已知的一般性结论。演绎推理所得到的结论，仍应由实验加以验证，或经过大量的实践验证之后，才能正式成立。

力学中的冲量定理、动能定理导出，热学中熵的定量描述，电磁学中位移电流、光的电磁说提出，光学中惠更斯-菲涅耳原理的定量描述等，都是使用了数学演绎的方法。

## 2.7 分析与综合

在研究和解决物理问题的过程中，分析方法与综合方法往往同时应用、互相补充。因此，可以将分析方法与综合方法作为一个整体进行考察。比如，力学中的隔离体受力分析就是典型分析方法，物理规律的研究必须结合综合方法。

### 2.7.1 分析

**1. 分析的定义**

分析方法就是把研究对象分解为各个组成部分和要素，然后分别加以研究，从而揭示事物的属性和本质的方法。

分析方法是以事物的整体与部分的关系为客观基础的。任何事物都是由各个部分或要素组成的，事物整体的各种属性、关系等，必然从事物的组成部分和要素的运动及相互作用、相互联系中表现出来。因此，人们为了从总体上揭示和把握事物的性质以及运动规律，就必须了解其各个组成部分和要素的性质、特点和相互联系。客观事物的整体与部分的这种关系，使运用分析方法解决客观事物的许多问题不仅成为可能，而且成为现实。例如，在研究力学现象时，为了分析物体的受力情况通常采用“隔离法”，就是把研究对象作为部分从系统整体中隔离出来，变整体为部分，变复杂为简单，化难为易，这就是分析方法。

**2. 分析方法的类型**

在运用分析方法时，由于科学研究对象的不同和内部矛盾的不同因而分析方法也就有所不同。适用于各门学科的基本分析方法大体有定性分析法、定量分析法、因果分析法、可逆分析法、系统分析法、结构分析法、比较分析法、分类分析法和数学分析法等。

定性分析法是为了确定研究对象是否具有某种性质时所采用的方法，主要解决“有没有”“是不是”的问题。例如，在研究阴极射线时，可以利用带电粒子在磁场中运动受到洛伦兹力而发生偏转的性质，来定性地分析阴极射线所带电荷的性质，为以后定量分析阴极射线提供依据。在实验分析中，定性分析是粗浅的、精度不高的分析，但它能为定量分析提供条件并指明方向。

定量分析是为了确定客观对象各种成分的数量的方法，主要解决“有多少”的问题。在实验分析中，定量分析要求精度高，操作要正确谨慎；在抽象理论思维分析中，定量分析主要从物质量的方面进行分析，且往往与数学方法相结合。经过定量分析，可以找出事物量的方面的运动规律。例如，氢的同位素氘就是尤莱在1932年对原子光谱的定量分析中通过里德伯常数的变化发现的。

因果分析法是为了确定引起某一现象变化原因的分析方法，主要是解决“为什么”。如原子的核式结构模型就是采用这种方法确定的。

可逆分析法，即分析作为结果的某一现象是否又反过来作为原因，从而产生本是原因的那一现象，这是一种逆向思维的方法。例如，奥斯特实验说明电流可以产生磁场；法拉第逆向分析是否可以由磁场来产生电流。据此，法拉第通过实验在1831年概括出电磁感应定律。

数学分析法，是现代物理学最常用的分析方法，它使物理学发展到了精密化和完善化的阶段，其中有一种特殊的分析方法叫元抽象法或元过程分析法，其特点是从某种物理现象中抽取任意一小部分进行研究。例如，从一定质量分布的刚体内“分离”出一个非常小的质量元，从流体内“抽取”一个非常小的体积元，从连续带电体中“分割”出一个很小的电荷元，从连续变化的物理过程抽取一个元过程等，分析这个小单元的特点并描述它的各物理量间的相互关系和变化规律，进而建立描述整个物理系统的微分元。由此求出物理系统在某一特定条件(定解条件)下的瞬时状态，继而把整个物理过程的运动特点和趋势表述出来。元过程分析法是用数学工具研究物理系统常用的方法，它属于定量分析的范畴，是辩证法在物理学研究中的具体体现。

**3. 分析方法在物理学研究中的作用**

分析方法作为重要的方法论手段，在物理学研究中具有十分重要的意义。第一，通过分析方法，可以将复杂的问题分解为简单的问题进行研究。这样既可以大大简化研究工作，又可以更深刻地揭露研究对象的本质。第二，通过分析方法，使研究工作由表及里地逐步深入到问题的核心，才会有新发现。物理现象的特性和规律往往都比较复杂，而且隐藏得比较深，只有使用分析方法，才能将各种表面现象从假象一层一层地剥去，才能发现事物的本来面目。第三，运用分析方法，可以弥补归纳法的不足。归纳法是在大量经验材料的基础上进行概括，决定了归纳的结论是一种不够严密的带或然性的推理。它虽然深入到事物和实验中去，但是，也不能保证通过感性认识就能全面、正确把握、反映研究对象。还必须借助于分析方法进行理性分析才能得出正确的结论。

**4. 分析方法的局限性**

分析方法也有其局限性。主要表现在：第一，分析方法尽管可以对事物进行必要的分割，以达到孤立地分析研究，从而把人们对物理世界的认识引向深入，但往往可能将人们的眼光限制在片面、狭窄的领域里。第二，尽管从研究的需要出发，分析方法必须把本来相互联系的有机整体暂时分离开来，以便从相对静止的角度进行单独的观察和研究，然而久而久之，也容易使人们忽视事物之间的有机联系，养成一种孤立、静止和片面地看问题的习惯。分析方法的这些局限性，要求物理研究工作者对所研究的物理对象必须做到具体问题具体分析，不能使对物理对象的认识仅仅停留在分析阶段上，还应该继续去把握整体与部分的关系，进行进一步的综合。

## 2.7.2 综合方法

**1. 综合方法的定义**

综合方法就是把研究对象的各个部分、方面和因素联合起来加以研究，从而在整体上把握事物的本质和规律的思维方法。但是这种联合不是机械的联合，而是建立在辩证关系基础上的联合。目的在于寻找它们的联系，弄清它们之间的相互作用和影响，从本质上认识它们的同一性，然后再加以概括和提炼。因此综合必须以比较作自己的前导，只有进行充分的、全面的比较，才能深刻地认识同一性的物理内容。这就要求人们首先把握各部分、各层次、各要素的特性以及作用和规律，这是综合方法的前提和准备。其次，全面掌握事物各部分、各方面的特点及其内在联系，并通过概括和升华，以事物各部分、各种属性和关系的真实联结和本来面目来复现事物的整体，综合为多样性的统一。一句话，综合方法变简单为复杂，变分离为统一，变局部为整体。综合方法在物理学研究中的任务和目的，是揭示和把握物理对象的根本性质和根本规律，解决物理实验和理论方面的大量研究课题。

**2. 综合方法的分类**

综合方法的形式很多。从综合方法的规模来说有小型综合和大型综合，从具体的方法而言，有对称法、移植法和系统法等。

小型综合主要指某一具体课题、具体研究对象的综合，它是以具体对象与课题的分析作为基础，全面地概括出研究对象与课题的分析作为基础，概括出研究对象的特性和规律。例如盖·吕萨克(Joseph Louis Gay-Lussac，1778—1850)在对气体各状态进行分析研究后，又进行了综合，阐明了气体运动的基本特性和规律。可见，小型综合就是具体综合的方法，它可以是直观模型的综合、原理的综合或数学模型的综合。

大型综合往往牵涉多个分支领域，要概括数代人的工作，它是对已知的理论和实践的大总结，并在新的基础上以更简明的形式，更深刻、更本质、更全面地把握物理运动的特性的规律。这样的综合在物理学发展史上往往有特别重要的地位，它对整个科学技术的进步和对社会生产的发展，都有着巨大的推动作用。物理学曾经历了五次大型的综合，即牛顿力学的建立，热力学和统计物理的建立，麦克斯韦电磁场理论的建立，相对论的建立，量子力学的建立。现在，物理学家正在酝酿更大范围内的综合，即统一场论的研究。

对称综合，就是从物理世界的多样性出发，综合各种物理客体之间的对称性联系，以探索物理研究对象之间的内在本质和规律的思维方法。对称现象是辩证法的生动体现。物理学家根据世界的对称性综合各种现象，通过预言、设想来推测未知事物的存在，提出各种各样的假说。例如，自从1897年发现电子后，狄拉克综合前人的研究结果，预言了正电子的存在。1932年，美国物理学家安德孙(Carl D. Anderson，1905—1991)在宇宙射线实验中发现了正电子。由此开始，许多物理学家根据对称性的特点掀起了寻找与已有粒子相对称的反粒子热潮。理论物理及实验都证明，已发现的所有基本粒子都有与之相对称的反粒子存在，组成基本粒子的各种夸克也都有自己的反夸克。由于整个世界的物质运动是处于对称和不对称的矛盾之中，因此，在物理学研究中，充分运用这种方法对于构造物理学理论有着十分重要的作用。

**3. 综合方法在物理学研究中的作用**

(1) 运用综合方法可以导致物理学的新发现，推动理论的发展

综合方法在物理学发展史中占有极其重要的地位。当感性经验以及对某一物理现象或过程的某些方面的认识积累到一定程度时，物理学家就能统观全局，进入整理材料、综合研

究的新时期,从而做出重大的发现和发明,建立物理学理论或提出重要的科学假说。如爱因斯坦在分析光电效应实验规律的基础上进行了高度综合,以光量子假设给出了光电效应的圆满解释。再比如,1932 年发现中子后,苏联物理学家朗道(Lendau,Lev Davidovich 1908—1968)综合了天文学的研究资料,提出宇宙中存在一种密度极高的中子星的假说。两年后,美国的伯德及茨威斯基也发表了他们的看法。1968 年,人们果然从蟹状星的中心找到了这种星。物理学史表明,物理假说、物理定律的提出,以致物理学理论体系的建立都常常需要对物理现象各方面的认识进行综合,对许多人的研究成果进行综合,通过综合建立物理学理论,导致科学发现。牛顿是创立经典力学的杰出代表,然而他的万有引力定律和力学三大定律并非完全是他本人的匠心独运,而是对前人关于力学和天体力学的研究成果加以创造性综合的结果。牛顿的成功既是综合的结果,也是科学发现的结果,是由综合而导致发现的突出典型。

(2) 在物理学研究中,运用综合方法可以克服分析方法的局限性,弥补演绎法的不足

分析是人们探索自然奥秘的过程中所采取的一种手段和最初环节,是为综合做准备。在物理学研究中,对某一物理对象进行科学分析时,时时想到综合就不至于使人的眼光停留在分析阶段上,就能克服分析方法的不足,以便从整体上把握事物的内在联系、本质和规律,克服片面性。演绎法是从一般推导个别的思维方法,它无法反映具体事物属性的多样性。而综合是在分析研究由实践积累起来的丰富而真实的具体经验材料的基础上进行的,使得出的一般结论能够反映出研究对象的多样性本质。离开了综合,只靠演绎,则很容易误入歧途。因此可以说,综合方法弥补了演绎法的不足。

## 2.7.3 分析与综合的关系

### 1. 分析与综合的辩证统一

分析方法的关键在“分”,而综合方法的关键在“合”,在具体运用的过程中,两者的步骤做法是完全相反的。虽然其他方法如比较和分类、归纳和演绎等都有对立的一面,但是,远没有分析和综合的对立那么尖锐和鲜明。另一方面,由于分析方法和综合方法的作用和局限性,又决定了它们之间存在着其他方法所不能比拟的相互依存和相互转化的紧密联系。

分析方法反映了认识过程和规律的一个方面。人们对物理现象和过程的研究和认识,总是逐个部分、逐个方向、逐个层次地进行的。这也是一个由浅入深、由表面到本质、由知之不多到知之较多的过程。这个过程就需要分析方法,也只有分析方法才能完成。但是,经历了这个过程,人们仍未能全面地把握住研究对象,只是零碎地、孤立地认识研究对象某些方面和部分。如果停留在这个阶段,充其量只是凑合一些材料,更有甚者,还可能将人的眼光限制在狭隘的领域里,把本来互相联系的东西割裂开来考察,就容易造成一种孤立的、片面的思维习惯。只有将各部分、各层次之间的联系和相互作用弄清楚,明确了它们在整个研究对象中各自的地位,才能从根本上把握研究对象,而这后一过程,正是综合方法要完成的使命。对于一个完整的认识来说,分析和综合是不可缺少的、相互补充的。运用综合方法可以统观全局,全面地考察对象,从整体上把握事物,避免了单纯使用分析方法的弊端;而综合方法又必须以分析方法为基础,只有分析得细致、透彻,才能准确全面地综合,也才能揭示整个研究对象的性质和规律。离开分析就不可能进行综合。总之,在物理学研究中,不管自觉不自觉,实际上都是把分析和综合作整体方法使用的,没有任何方法比分析和综合配合得更密切的了。

分析研究对象必有一个指导思想,这个思想本身又是综合的结果。它是研究者把握已

知的理论和实践，根据研究对象的具体情况综合的产物。正是在这个思想指导下，才能进行具体的分析工作。因此，分析方法的使用，又是以综合作为前提的。当然，这种综合是另一种含义和内容的综合。如果没有这个综合，就不会有全面的分析。这一方面也显示了分析和综合的密切联系。

分析方法和综合方法的联系，还体现在研究对象地位的转化。这种转化在物理学研究中表现得尤其明显。物理现象和过程尽管错综复杂地交织在一起，但是，各现象的层次还是清楚的。同一现象和过程，在某一层次上，它可能是综合研究的对象，但是，在另一层次上又可能变成分析研究的对象。这也从一个侧面体现了分析和综合的统一。例如，人们以原子结构为研究对象，目的在于把握原子的特性和规律。对它的研究，就分子结构这一层次来说，是分析研究的一部分，因为它能从组成上说明分子的特性。但是，就原子核和电子这一层次来说，它是综合研究，因为原子的特性和规律，如原子量、原子序数、化合价等都是原子内部的电子和原子核相互作用的整体效应，必须从原子内部的相互联结上和总体上来把握它。

**2. 分析与综合的应用类型**

在物理思维中，分析与综合方法主要有如下几种：

(1)“从复合到单一，从单一到复合”的分析与综合的方法

在对复杂的物理对象、物理现象和物理过程的研究中，我们往往把研究对象以及所产生的物理现象分解为许多单一部分或单一过程进行研究，然后，把分解的各部分再结合成整体来认识。这种方法称为“从复合到单一，从单一到复合”的分析与综合的方法。

例如，在力学中所采用的“隔离体法”就是“从复合到单一，从单一到复合”的分析与综合的方法，这种方法是先想象从连接体中隔离出所要研究的某一物体（或物体的某一部分）。然后单独分析“隔离体”的受力情况，结合“隔离体”的运动情况，全面考察周围物体对它的作用，列出“隔离体”的运动方程，得出“隔离体”的运动规律。最后，考察各“隔离体”的相互关系，得出整个研究对象的运动规律。

应当指出，运用“隔离体法”解决力学问题要用联系的观点，把研究对象隔离开来孤立地研究，正是为了找出物体（系统）内部各部分之间的相互联系及相互作用。例如，要求出连接体（或重叠体）各部分间的相互作用力就必须用“隔离体法”，只有把连接体各部分分隔开来，分别以各个“隔离体”为研究对象，连接体各部分间相互作用的情况才能表现出来。用“隔离体法”不仅可以解决物体受力问题，而且还可以用来解决其他力学问题，关键在于正确确定研究对象——隔离体。

在研究复杂的振动和波动问题时，也要利用“从复合到单一，从单一到复合”的分析与综合的方法。在物理学的各分支中，广泛地存在着振动和波这两类周期现象，例如，弹性振子、声振动、声波、交流电流、电磁振荡、电磁波等，这些振动和波往往是很复杂的，但它们可以分解为一系列各种频率的谐振动（由正弦或余弦函数所描写的振动）和谐波（由正弦或余弦函数所描写的波）的叠加，我们可以先对这些谐振动和谐波进行研究，然后再将它们叠加起来，综合得出复杂振动和波的规律。

对曲线运动的研究，采用了分解与合成的方法，先把实际运动分解成两种或三种相互独立的单一的分运动，应用力学规律逐个进行研究，然后再综合起来，得出曲线运动的规律。

总之，“从复合到单一，从单一到复合”的分析与综合的方法是物理学研究、处理问题时常用的思维方法。

(2)“从整过程到分过程,从分过程到整过程”的分析与综合的方法

物理学中不仅存在许多复合问题,而且存在许多由几个分过程组成一个物理过程的复杂问题。为了研究这些复杂的过程,我们往往将之分解为几个简单的分过程来处理。先搞清每一个分过程的规律,然后再把这些分过程联系起来,得出整个过程的规律。这种方法称为“从整过程到分过程,从分过程到整过程”的分析与综合的方法。许多物理问题特别是力学问题常常可用这种方法来解决。

例如,火车从甲站到乙站,假设轨道是平直的,火车进站和离站的速度都是均匀变化的,对这个过程我们也必须分解成三个不同的分过程来处理,即第一分过程——火车离站时的匀加速直线运动;第二分过程——火车中途的匀速直线运动;第三分过程——火车进站时的匀减速直线运动。在研究这些简单分过程的基础上,得出火车在整个过程中的运动规律。

又如,在热学中,卡诺对热机的研究,也利用了这种分析与综合的方法。他把理想热机中的热力学过程分解为前后相随的四个分过程,先研究每一分过程的特点及规律,然后研究这些过程的相互联结,终于“彻底”研究了著名的卡诺循环,从而得出了热机效率的一般公式。

(3)“先分析,后综合”的分析与综合的方法

在物理学中,用“控制条件”来探索物理问题和物理规律的方法,即为“先分析,后综合”的思维方法。

在热学中探索气体状态变化的规律时,就运用了“先分析,后综合”的思维方法。气体的状态涉及压强、体积和温度三个参量,为了得到气体状态变化时这三个参量之间的关系,可以采用“控制条件”的实验方法。先保持一个参量不变,研究其他两个参量之间的关系,分别得出在不同条件下气体状态变化的三个实验定律,然后再综合总结压强、体积和温度三个参量之间的变化规律,推出理想气体的状态方程,这就是所谓的“控制变量法”。

物理学中很多重要的实验规律,如牛顿第二定律、欧姆定律、法拉第电磁感应定律等在建立过程中都应用了“先分析,后综合”的方法。另外,从物理学某一分支理论的形成来讲,先是分析某一问题,得出局部范围内的物理规律;然后将这些规律综合起来,得出反映物理事物本质的一般的物理理论。在物理学发展的整个历史过程中,作为认识发展过程中的不同阶段,往往表现为先以分析为主,后以综合为主。

## 2.8 假说

假说是以物理事实和科学知识为根据的猜想。假说即是指科学的猜测或设想。假说经过实践的检验,可以转化为理论。理论随着实践的发展又将接受新的假设的挑战。创造条件促使假设向理论的转化,将推动科学认识的向前发展,这就是假说的作用。

### 2.8.1 什么是假说

从实践方面看,人们会碰到许多根据已有知识难于解释的问题,于是进行各种猜测。有的猜测是没有事实和科学根据的臆想;有的是以事实和科学知识为根据的猜想。假说是指科学的猜测或设想。

从科学思维的角度看,假说是一种复杂的理论思维形式,是人们运用科学思维,根据一定数量的事实材料和已有的科学理论,对未知的事物及其规律所作的推断和假定,是一种带有推测性和假定性的理论形态,是没有经过实践充分证实的理论。

从科学研究方法角度,一切科学无不是经过假说而发展起来的。历史上新理论的产生,总

是以假说的形式出现,科学的假说都是以一定的事实为基础,不是毫无根据地瞎说或臆测,同时,它有待于实践的进一步验证,所以不是一种完全肯定的结论。当物理假说被实验所证实或被理论所证明的时候,它就变为理论,变为知识,就不再是假说了,如分子运动论、日心说、光的电磁说与光的粒子说、卢瑟福的原子核式结构以及德布罗意的物质波等都曾是物理假说。

假说的定义:假说是指按照预先设定,对某种现象进行的解释,即根据已知的科学事实和科学原理,对所研究的自然现象及其规律性提出的推测和说明,而且数据经过详细的分类、归纳与分析,得到一个暂时性但是可以被接受的解释。任何一种科学理论在未得到实验确证之前表现为假设学说或假说。有的假设还没有完全被科学方法所证明,也没有被任何一种科学方法所否定,但能够产生深远的影响。

### 2.8.2　假说的特征

**1. 假说具有猜测性**

假说之所以称为假说,就是因为它是一种“毛坯”,是具有一定猜测性的理论“预制品”,在未证实之前只能说是对自然现象及其规律的推断、猜测。假说只有通过实践检验和证明,才能上升为理论。例如,哥白尼的“日心说”、康德(Immanuel Kant,1724—1804)——拉普拉斯(Laplace,Pierre-Simon,1749—1827)的“星云假说”“大爆炸宇宙说”“黑洞假说”等。近代物理中卢瑟福核式结构模型、玻尔量子假说等,在未证实之前均为假说,不能看成是一种完美的理论。有的假说至今未上升为理论,因为,其未得到实验的验证。

**2. 假说具有科学性**

假说虽然是一种想象或猜测,但它的提出不但要以实验材料与经验事实为基础,而且要以现有科学理论为依据。所以假说不是无根据的猜测和幻想,不是主观臆造。假说一旦失去事实基础和科学依据,又未经受一定的实践或检验,它也就失去了存在的价值。卢瑟福的原子核式结构模型,虽然有其很大的缺陷,但是它有科学依据,又能经得起 $\alpha$ 粒子散射实验的检验,能推算出原子和原子核的半径。玻尔的原子模型是在卢瑟福模型的基础上发展起来的,它克服了卢瑟福模型的某些缺陷,引入了量子化的概念,因而是更为合理的一种新的假说,但是它还是有自己的缺陷。因此每一种假说都有科学依据,但它又随科学的不断发展而“进化”,并且等待接受实验、实践检验和证明。

**3. 假说具有可变性**

假说是一种对实践证实的东西的猜测,因而通过实践检验可能成为真理而发展成一种理论,也可能成为谬误而被淘汰,也有可能被证实具有某种不完整性而发展成为一种新的假说。有时也通过相反的假说之间对峙和争论,形成一种变动更迭、新旧交替的局面,使假说得以发展,一步步逼近真理。例如,关于太阳系演化的假说,18 世纪康德-拉普拉斯提出了“星云假说”,到 20 世纪 70 年代又出现了“星子假说”“陨星假说”“宇宙大爆炸说”等,这些假说都有科学价值,因为都从某一角度反映了太阳系起源的可能性,都为解决太阳系起源作了贡献。关于光的本性的认识,历史上许多科学家曾经先后提出了对立的不同假说。牛顿对光的反射、折射、色散等现象进行了大量的研究,认为光是由发光体射出的微粒所组成,后人称为“光的微粒说”,但“微粒说”无法解释光的衍射和干涉现象。惠更斯在长达 15 年的时间里,研究了光的反射、折射、衍射等现象,认为光和声现象相似,是在一种特殊弹性物质(以太)中传播的弹性机械波动,他所建立的“惠更斯原理”是“光的波动说”的核心。麦克斯韦在研究电磁场理论的基础上推算出电磁波的传播速度恰等于光速,由此推断光也是一种电磁波。爱因斯坦在研究光电效应时,把普朗克提出的辐射能量不连续——量子化概念推广到

辐射的传播过程中,提出光量子假说,认为光辐射也是以一束一束能量形式存在的,运用光量子假说,成功地解释了光电效应。光的“微粒说”“波动说”“电磁说”“光子说”的先后提出,反映了人们从不同侧面对光的本性认识的逐步深入,而且新提出的假说总比原先的假说更接近事实本身。从光既具有粒子性又具有波动性的事实,人们逐渐认识到光具有波粒二象性。又如“以太”假说在科学事实面前证明是一种谬误,因而被人们抛弃。

另外,假说还具有预见性,它是对事物的本质、事物的内在联系、事物的规律性的猜测和推断,已具有一定的预见性。当然,这种预见性不一定准确,也表现出多样性。在科学研究中,对同一现象及其规律可以做出两种或两种不同或多种不同的假说,以供比较研究。一个已形成的假说应包括两部分:一是假说的核心部分,另一个是假说的推论部分。假说还具有可验证性,无法验证的,不能称其为假说。

### 2.8.3 假说的形成过程

**1. 提出假说阶段**

在这个阶段里,研究者首先通过观察和实验等方法广泛地搜集与被研究现象有关的事实材料。然后运用有关知识,对已有材料进行逻辑分析,并通过一定的推理形式提出假说。

例如,魏格纳在提出“大陆漂移说”时,首先把注意力集中在搜集事实及对事实的分析上。他在《海陆的起源》中说:“任何人观察南大西洋的两对岸,一定会被巴西与非洲间海岸线轮廓的相似所吸引住。不仅圣罗克角附近巴西海岸的大直角凸出和喀麦隆附近非洲海岸线的凹进完全吻合,而且自此以南一带,巴西海岸的每一凸出部分都和非洲海岸的每一个同样形状的海湾相呼应。反之,巴西海岸有一个海湾,非洲方面就有一个相应的凸出部分。如果用测量工具在地球仪上测量一下,就可以看到双方的大小都是准确一致的。”

然后,研究者在掌握上述事实的基础上,根据牛顿力学的某些原理对事实进行分析,设想远古时代地球上只有一整块陆地,称泛大陆,在它周围是一片广阔的海洋。后来由于天体的引力、潮汐和地球自转所产生的离心力,使原大陆分成若干小块。同时,又将地球表面的状况与冰川漂移的情况进行类比,联想到较轻的刚性的大陆块是漂浮在地壳内较重的黏性液体——岩浆之上的。它们在分成小块以后,就像漂浮的冰山一样逐步远离开去。美洲脱离了欧洲和非洲向西移动,越漂越远,在它们之间就形成了大西洋。

又如,达尔文在《物种起源》出版前22年,就开始搜集并且分析了动植物的变种,把自然界中的物种与家养动物的变种作了类比,初步觉察到自然选择的力量。然后他又继续研究,归纳了在随贝格尔号航行中搜集到的大量动植物变种的材料,得出了物种不是不变的结论,提出了生物进化学说。

**2. 验证和确立假说阶段**

在此阶段中,研究者要集中精力对假说进行逻辑推演,并根据理论和事实进一步论证假说。首先,从已假定的初步原因出发,推导出一系列的相应结果。其次,验证由假设所推导出结果是否符合客观现实和已有的理论。然后,根据验证的结果决定对假说的取舍。如果假说被证实,则假说就可以确立;如果假说不能得到证实,则要对假说提出补充、修改,或者另立假说。

例如,根据几何学、考古学等基本理论,从“大陆漂移说”的基本观点出发,可推导引申出以下结果:各大陆都可以像拼板玩具那样完全拼合起来,而且各大陆块边缘之间的吻合程度应当是很高的;大西洋两岸的动植物化石也应当是相同的;大西洋两岸的距离应在继续增大。经验证,上述情况与事实基本相符,因而,“大陆漂移说”被确立下来。

再如，达尔文提出了生物进化的基本思想以后，还继续做了大量工作，他亲自做实验，收集有关实验资料，并运用生物的地理分布、地质史、生物的亲缘关系等方面的观点和材料，证明了自己理论观点的正确性。假说只有经过这样比较充分的研究工作，才算是最后完成。

一般来说，在提出假说阶段，主要运用归纳推理和类比推理；在验证和确立假说阶段，主要运用演绎推理。

## 2.8.4　假说在科学发展中的作用

假说作为一种重要的科学研究方法，在物理学中具有其他方法所不可替代的作用。其作用可以从图 2-8-1 中看出。

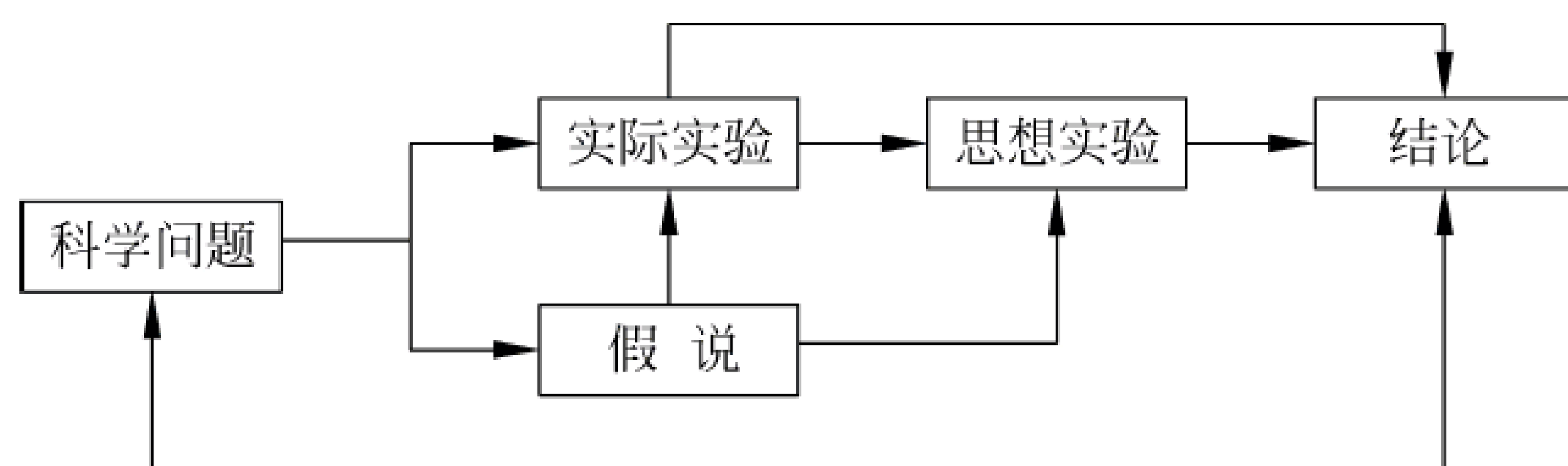



图 2-8-1　假说在科学发展中的作用

**1. 假说是科学问题过渡到科学理论的桥梁**

一个自然现象，在其未被揭示出科学本质之前，人们对它的认识是很不完整的，甚至是片面的，只能借助于假说的形式进行研究与探索。当某一假说被大量事实所证实时，它就发展成一种理论；当新的科学事实又积累到一定程度与假说相矛盾时，又必须提出新的假说或修改、补充原来的假说，以便能圆满地解释事实，进而促进理论的进一步研究与发展。因此假说是物理学研究中理论发展必不可少的方法与桥梁。例如，从黑体辐射问题的研究中出现的"紫外灾难"到 1900 年普朗克的量子假说，到 1905 年爱因斯坦"光量子假说"，到玻尔量子理论解释原子问题建立旧量子论，到海森伯、薛定谔提出量子力学，再应用爱因斯坦相对论提出相对论量子力学的整个量子理论的发展，无不体现了从假说—理论—新假说—新理论的循环发展模式，而每一次的发展都是对前一层次理论（假说）的继承、完善和修改，又是后一层次理论（假说）的重要台阶。因此，只要物理学发展着，假说便是永远不可缺少的一座桥梁。

**2. 假说对物理观察和实验具有先导作用**

假说在科学研究中具有一定的猜测性。但不是盲目的无目的的猜测，而是科学的预测、有计划的研究。有了假说便有了有计划、有目的的实验和观察。因此假说对物理观察和实验具有先导作用。例如在天体物理研究中，1845 年，法国天文学家勒维烈（Le Verrier，Urbain Jean Joseph，1811—1877）和英国天文学家亚当斯（John Adams，1735—1826）根据万有引力定律计算的天王星轨道的偏离值，预言有一颗新的行星存在。1846 年 9 月 23 日，由柏林天文学家加勒（Johann Gottfried Galle，1812—1910）在勒维烈预言的位置偏离 1°的地方果然找到了这颗新行星，并命名为海王星。又如 1919 年英国两个科学观测队按爱因斯坦关于光线通过太阳表面发生偏转的理论计算值所做出的观察，都是有计划的实验活动。

## 2.8.5　假说的应用领域

假说的产生是有其客观根据的，一般假说方法可以应用于以下场合。

**1. 为实践开辟新的研究领域**

在这些研究领域中，科学知识还未接触过或者接触很少，必须通过创立假说才能对其中有

限的科学事实给予概括和解释。例如,人们根据宇宙飞船所提供的资料,开始研究星际空间以及人在星际旅行中所遇到的各种问题,提出了关于人体在超重和失重状态下的生理问题的假说和关于其他行星上的植物的假说,前者是"宇宙医学"的雏形,后者是"天文植物学"的雏形。

**2. 原有的理论与事实发生矛盾**

随着人类社会实践的发展,当原有的理论与事实发生矛盾时,传统观念就发生动摇,这就必须建立另一种假说去解释新事实。例如"燃素说"认为金属加热后重量增加是有许多"燃素"跑到金属里面去的结果,后来精密实验表明,金属加热后重量仍不变,表明没有什么"燃素"跑到里面去,这就与"燃素说"发生了矛盾,对热的本质必须用新的假说来说明,新假说认为热的本质是物质粒子的运动。

**3. 原有理论体系不够完善**

如果原有理论体系不够完善,存在某方面的缺陷,就要通过假说的形式去完善它。

总之,假说的产生有其客观的必然性,不是从人的愿望出发任意制定出来的。假说是理论的过渡形态,因此科学理论本身的特性与功能,就成为提出假说时应考虑的原则:解释性原则、对应性原则、简单性原则、可检验性原则。

科学假说一旦经受住了实践检验,具有了解释性和预见性,就将转化为一个新的科学理论。但在实际的研究中,我们常常遇到的情况是,科学假说与科学理论之间并没有一条明确的界限,从哲学的原则上看,二者的区分就在于实践。实践检验一方面使科学假说上升到理论;另一方面,又使理论在新的事实面前成为科学假说。理论或者能够成功地解释新事实,扩大自己的应用范围;或者在新事实面前发现自己的局限性,让位于新的科学假说。科学史表明,任何理论都是相对真理,都要在认识真理的长河中为新的科学假说所代替,任何理论都保留着某种假说的痕迹。

## 2.9 数学方法

物理学的发展与数学密不可分,数学对于物理不仅是一种计算工具,也是物理学的思维工具,只有具备较好的数学基础才能更好地掌握物理概念和定律,才能深入地揭示和描述物理现象、物理过程的实质。数学方法为物理学研究提供简明精确的形式化语言、数量分析、计算方法和逻辑推理等内容。数学方法把物理规律、物理图像等与数学形式紧密联系起来。英国著名的物理学家汤姆说过:你所研究的问题,如果不能用数学表示,那么你的认识是不够的,不能令人满意的,可能只是初步的认识,在你的思想上,还没有上升到科学的阶段。另一方面,物理学的发展也对数学提出了新的课题,从而也促进了数学的进一步发展。

关于高等数学在物理学理论建立中作用的详细讨论,请读者参阅本书第12章。

### 2.9.1 数学的本质与特点

**1. 本质**

各种运动形式在空间形式及其数量关系的概括。

**2. 特点**

(1) 高度的抽象性

如虚功原理

$$\delta W=\sum_{i=1}^{n}\boldsymbol{F}_i\cdot\delta\boldsymbol{r}_i=0$$

（2）逻辑上的严密性

如麦克斯韦方程组

a. 四个积分方程

$$\begin{cases} \oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{S} = \dfrac{q_0}{\varepsilon_0} \\ \oint \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{S} = 0 \\ \oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\dfrac{\mathrm{d}\phi}{\mathrm{d}t} \\ \oint \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} = \mu_0 I \end{cases}$$

b. 四个微分方程

$$\begin{cases} \nabla \cdot \boldsymbol{E} = \dfrac{\rho}{\varepsilon_0} & \leftarrow \text{有源场} \\ \nabla \cdot \boldsymbol{B} = 0 & \leftarrow \text{无源场} \\ \nabla \times \boldsymbol{E} = -\dfrac{\partial \boldsymbol{B}}{\partial t} & \leftarrow \text{变化的磁场产生电场} \\ \nabla \times \boldsymbol{B} = \mu_0 \varepsilon_0 \dfrac{\partial \boldsymbol{E}}{\partial t} & \leftarrow \text{变化的电场产生磁场} \end{cases}$$

（3）数量上的精密性

如波动动力学方程，加上初值条件和规范的边界条件，可以严格求解

$$\begin{cases} \nabla^2 \boldsymbol{E} - \mu_0 \varepsilon_0 \dfrac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0 \\ \nabla^2 \boldsymbol{B} - \mu_0 \varepsilon_0 \dfrac{\partial^2 \boldsymbol{B}}{\partial t^2} = 0 \end{cases}$$

（4）深刻上的辨证性

如极限——曲与直、微分——无穷小……

增量趋于零→ $\lim\limits_{\Delta x \to 0} \dfrac{\Delta \boldsymbol{y}}{\Delta x} = \dfrac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}x} = \boldsymbol{v}$ ←无穷小的比值发生了质变

## 2.9.2　数学是物理学的表述形式

**1. 经典力学的例子**

牛顿三定律：$\boldsymbol{v} = \boldsymbol{0}$；$\boldsymbol{F} = \dfrac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t}$；$\boldsymbol{F} = -\boldsymbol{F}'$

万有引力定律：$\boldsymbol{F} = -G\dfrac{m_1 m_2}{r_{12}^2}\boldsymbol{r}_0$

胡克定律：$\boldsymbol{F} = -k\boldsymbol{x}$

摩擦定律：$\boldsymbol{F} = \mu \boldsymbol{N}$

**2. 经典热力学的例子**

热力学第一定律：$\mathrm{d}Q = \mathrm{d}W + \mathrm{d}E$

热力学第二定律：$S = k\ln W$

**3. 经典电动力学的例子**

麦克斯韦方程组 $\begin{cases} \oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{S} = \dfrac{q_0}{\varepsilon_0} \\ \oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\dfrac{\mathrm{d}\phi}{\mathrm{d}t} \\ \oint \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} = \mu_0 I \\ \oint \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{S} = 0 \end{cases}$；物质方程 $\begin{cases} \boldsymbol{D} = \varepsilon \boldsymbol{E} \\ \boldsymbol{B} = \mu \boldsymbol{H} \\ \boldsymbol{J} = \sigma \boldsymbol{E} \end{cases}$

**4. 经典光学的例子**

矢量的线性叠加原理；麦克斯韦方程组；惠更斯-菲涅耳原理。

### 2.9.3 数学是创立和发展物理学理论的主要工具

物理概念、原理、定律以及物理理论的建立都离不开数学工具。

我们知道，一个物理理论建立之前，必须具有自己的概念，而物理概念的形成往往借助于数学，特别是现代物理学，它的内容越来越抽象，如果不借助数学工具，就很难概括。同样，物理原理、定律的建立也是如此，在经过量的测量之后，再建立量与量之间的联系——数学关系式。一些不是直接从实验中概括和抽象出来的物理理论，在创立它们的过程中，数学工具所起的作用就更加明显，比如，麦克斯韦的位移电流假说理论、量子力学的建立等就足以说明这一点。

物理学可以在数学形式中得到启发，从而促进物理理论体系的发展。由于数学也经常走在物理的前面，所以，为物理学的发展提供了极大的方便。例如，数学中的复数理论体系对后来的电工学、流体力学、振动理论的发展起到了重要作用。

物理学理论的研究从形式上来说也是一种数学研究，物理学对数学的依赖性在于数学的逻辑联系不是孤立的东西，而是物质运动在某一层次上的特性和规律。当人们研究、探讨这些特性和规律时，就必须使用数学工具才能认识和掌握它，也只有用相应的数学工具才能表述它。

### 2.9.4 物理学理论的应用要借助数学工具

在物理学研究中，从已知的理论出发，经过逻辑推理预见某些物理现象或规律，是常用的方法。这种预见往往都要借助于数学工具才能完成。质能关系式是爱因斯坦在相对论的基础上借助数学工具推导出来的。在完成相对论后，他发现物体的质量是它所含能量的量度。如果物体的能量为 $E$，那么其质量 $m$ 也就相应地为 $E/c^2$。他在 1905 年 9 月正式提出了这一重要发现，使当时无法解释的许多现象得到了完美的答案，比如，放射性元素镭为什么能够不断释放巨大能量的现象、太阳能的来源等问题，而且为原子核物理学、粒子物理学和原子能的利用奠定了基础。又如冲击波现象，首先是由非线性流体力学方程式推导出来的，后来才在实验中找到它的原形。物理学理论上的预见，基本上要借助于数学工具才能实现。

如同任何预见都要倾听实践的呼声一样，数学形式的物理预见也必须由实践检验。只有得到实践的认可，它才得到人们的承认，才能在物理理论体系中占据自己的地位。但是，物理理论的预见虽然也要由实验检验，有时某些依赖数学工具得出的物理理论预见被实验否决，却不能马上断定所用的数学工具是错误的。这是因为一个正确的物理预见不仅依赖于数学工具，而且依赖于其他因素，比如，是否正确把握具体的物理条件，其他方面的作用有没有全面考虑等。当然，也不能忽视在使用数学工具过程中的差错。为了求得正确的预见，必须保持数学演算、推导和证明的严密性，否则就会导致错误的结论。

物理学理论在工程技术中，有着广泛的应用。从日常的建筑到尖端的宇航技术，无不与物理理论相联系。在具体运用物理理论时，也要借用数学工具。可以这样理解，既然物理理论要依赖数学工具从现实原型中概括抽象出来，那么，将物理理论用到现实中去，实际上是一个逆的过程，这个过程也需要数学作为工具。

无线电技术，就是电磁理论的具体应用。从日常的广播、电视到复杂的遥控指挥系统，

它们的设计和应用都必须使用数学工具。火箭导弹技术也是物理学理论的具体应用,它牵涉到很多复杂的因素,例如,燃料的填装质量及消耗率、推动力大小的变化、结构质量及载质量、飞行的轨道,还有外界的条件如气象等因素的影响。要对这些因素加以综合,运用物理学理论进行处理,这本身就构成非常复杂的、大量的数学问题。不解决这些问题物理理论的应用就是一句空话。数学实际上是将抽象的物理理论同具体的工程应用联系起来的桥梁。

### 2.9.5　物理学促进数学的发展

从科学方法论的角度看,科学研究是从“问题”开始的,只有引出了问题,才能说开始了研究。而问题是由问题情境产生出来的,所以,创设问题情境对科学研究有“起点”的重要意义。物理学的问题就是问题情境的“创设”,即是数学应用,因而具有为数学研究创设问题情境的重要作用,也就是具有为数学思维活动即数学科学研究提供“起点”的重要作用。由问题情境得出有关的数学问题,使人们进行数学研究,促进了数学的发展。

例如,微积分的创立,首先是为了处理17世纪主要的科学问题的。当时的主要科学问题有四种类型,以第一类为例来分析。已知物体的运动距离表示为时间的函数的公式,求物体在任意时刻的速度和加速度;反过来,已知物体的加速度表示为时间的函数公式,求其距离和速度。这类问题是力学中研究运动时直接出现的(它们与人的实践有重要的联系)。从伽利略时代起,人们不仅利用数学来研究其他科学,尤其是力学和物理学,而且创立了数学和实验结合的科学方法论,数学成为科学特别是自然科学研究的首选方法,从那以后,数学在科学中的应用的需要就不断地推动着数学的发展。17世纪所涉及的速度和加速度每时每刻都在变化。计算瞬时速度,不能用早先计算平均速度的数学方法——用运动的时间除以运动的距离,因为在给定的瞬间,移动的距离和所用的时间都是0,而0÷0是无意义的。但是,按照力学规律,每一运动的物体在它运动的每一时刻必有速度,这是无疑的。已知速度公式求移动的距离问题,也遇到同样的困难,因为速度每时每刻都在变化,所以,不能用运动的时间乘任意时刻的速度,来得到移动的距离。因此,遇到的是一种缺少有效的数学工具的困难。由于处理这一类关于运动的问题是力学研究以及解决与之有关的实践问题(机械、武器设计等)的急切需要,这种急切的需要,促使人们把克服上述困难——寻找、创立新的数学工具,作为自己的任务,于是人们就进入了问题情境,从而对这类问题进行了深入的研究,提出了计算一个变量对另一个变量的变化率的问题及其逆问题,这就提出了数学问题。人们进入了数学研究过程,在这一研究过程及由其他物理问题(比如,光线通过透镜的折射和反射问题;求炮弹的最大射程、得到最大射程的最优发射角问题;行星的近日点、远日点问题;求重心、体积、面积等问题)促使人们在提出数学问题(比如,求切线的问题;求函数的极值问题;求曲线长度的问题)的研究过程中创立了微积分学。

### 2.9.6　数学物理方程

数学物理方程(简称数理方法)是指从物理问题中导出的反映客观物理量在各个地点、各个时刻之间相互制约关系的一些偏微分方程(包括一些常微分方程和积分方程)。

**1. 数理方程一般按照所代表的物理过程(或状态)可分为三类**

(1) 描述波动特征的波动方程

$$u_{tt}=a^2\Delta u+f \tag{2-9-1}$$

其中,$u=(x,y,z;t)$代表平衡时坐标为$(x,y,z)$的点在$t$时刻的位移(未知函数),$a$是波传播的速度,$f=f(x,y,z;t)$是与源有关的已知函数。

(2) 反映输运过程的扩散(或热传导)方程

$$u_t = D\Delta u + f \tag{2-9-2}$$

其中,$u=(x,y,z;\ t)$表示物质的浓度(或物体的温度),$D$ 是扩散(或热传导)系数,$f=f(x,y,z;\ t)$是与源有关的已知量,$u_t=\dfrac{\partial u}{\partial t}$。

(3) 描绘稳定过程(或状态)的泊松方程

$$\Delta u = -h \tag{2-9-3}$$

其中,$u=(x,y,z)$是表示稳定现象特征的物理量,如静电场中的电势等,$h=h(x,y,z)$是与源有关的已知量。

这三类方程,其未知函数 $u=(x,y,z)$的偏微商最高只有二阶,且各阶偏微商都是以线性的关系(即只有一次项)出现,故均属二阶线性偏微方程。在数学上,它们分别被称为双曲型方程、抛物型方程和椭圆型方程。

**2. 用数理方程研究物理问题的步骤**

数理方程是以物理学与工程技术中的具体问题作为研究对象的,简单地说,它是把对物理问题的研究"翻译"为对数学问题的研究。为了使这个"翻译"及其研究工作做得既完整又准确,一般须经过以下三个步骤:

1) 提出定解问题

它包括泛定方程,即数理方程本身和定解条件两方面,定解条件即确定具体问题的特解所需要的条件。同一类事物可用同一类方程描述,即泛定方程只是提供了解决问题的一般规律(同一类事物的共性),而定解条件提供了解决问题的具体条件(具体问题的个性),故它们总是同时提出,作为一个整体,称为定解问题。它是根据物理学的规律和实验资料而提出的。一般定解条件包括初始条件(初始位置等)、边界条件(描述物体的个性)和其他条件。

初始条件就是把体系在开始时($t=0$)的情况表达清楚。一般初始条件的个数等于方程对时间导数的阶数。对于波动方程而言,初始条件给出的是初始位移和初始速度;对于热传导(扩散)方程而言,初始条件给出的是初始温度(浓度)的分布;泊松方程和拉普拉斯方程不含时间变量,就不用给出初始条件。

边界条件就是把物体所处的环境表达清楚,一共分三类:

(1) 第一类边界条件:给出要求解的函数 $u$ 在边界上$(x_0,y_0,z_0)$的值 $f(x_0,y_0,z_0,t)$,则 $u(x,y,z,t)=f(x_0,y_0,z_0,t)$。

(2) 第二类边界条件:给出要求解的函数 $u$ 在边界上$(x_0,y_0,z_0)$的法向导数值 $f(x_0,y_0,z_0,t)$,则$\left.\dfrac{\partial u}{\partial n}\right|_{x_0,y_0,z_0}=f(x_0,y_0,z_0,t)$。

(3) 第三类边界条件:给出要求解的函数 $u$ 及其法向导数的一个线性组合在边界上$(x_0,y_0,z_0)$的值 $f(x_0,y_0,z_0,t)$,则$\left.\left(u+H\dfrac{\partial u}{\partial n}\right)\right|_{x_0,y_0,z_0}=f(x_0,y_0,z_0,t)$。

其他条件,有衔接条件和自然边界条件等。

2) 求解方法

提出了定解问题之后,实际上就完成了将物理问题"翻译"成数学语言的工作,下面紧接着面临的,当然应是对所提出的定解问题进行求解。数理方程的求解方法大致可归纳为如下几种。

(1) 行波法(或达朗贝尔解法)

行波法主要用于求解无界区域的齐次波动方程的定解问题。它是先求方程的通解,再

用初始条件去确定通解中的任意常数而得到特解。

(2) 分离变量法

分离变量法的主要思想是通过寻求齐次偏微分方程分离变量形式的解,将一个偏微分方程问题转化为多个常微分方程问题。运用它可以求解波动方程、输运方程和稳定场方程这三类方程。

用分离变量法解数理方程的步骤有:

① 将非齐次方程、非齐次边界条件齐次化;

② 分离变量(将解偏微分方程的问题转化为解常微分方程);

③ 求解本征值问题(确定本征值);

④ 将所求得的方程的解与本征值结合求得特解;

⑤ 特解叠加得通解;

⑥ 求系数。

(3) 积分变换法

积分变换法是通过积分变换把直接难于求解的偏微分方程、常微分方程或积分方程转化为易于求解的常微分方程或代数方程,它适用于求解无界区域或半无界区域的问题。但要求函数绝对可积。

用积分变换法解数理方法的步骤有:

① 对方程和定解条件(关于某个变量)取变换;

② 解变换后得到的像函数的常微分方程的定解问题;

③ 求像函数的逆变换(反演)即得到原定解问题的解。

这种解法的优点是减少了自变量个数,使偏微分方程化为了常微分方程求解,从而使问题大大简化。同时不必考虑边界条件的齐次与否,都采用一种固定的步骤求解,易于掌握。

(4) 格林函数(或积分公式)法

格林函数的基本思想是先得到点源的场,然后利用叠加原理求出在同样边界条件下任意源的场。

这种方法的优点在于:一旦求出了某个区域上的(狄氏)格林函数,便可“一劳永逸”地得到这个区域上的一切狄氏问题的解,而不管其具体的边界条件和方程的非齐次项的形势如何。

(5) 保角变换法

保角变换法的基本思想是利用解析函数所代表的变换的一些几何性质,把边值问题中区域的边界变简单,使之易于求解。

(6) 复变函数法

复变函数法是利用复变函数论中的调和函数与解析函数之间的关系,来求解二维调和方程的编制问题的方法。

(7) 变分法

变分法解数理方程的基本原理是:把一个微分方程的本征值问题或者定解问题和一个泛函的极值问题联系起来,使原来的方程是这个泛函的欧拉方程。然后用直接方法求出使该函数取极值的函数。由于这函数必定满足欧拉方程,它也是原方程的解。

此外,对于有些具体问题,当我们无法得到其解析解(或不需要得到其解析解)时,还可采用近似方法求解。

3）分析解答

用数理方程研究物理问题，仅仅求出了答案是不够的，还必须分析所求结果的物理含义并论证其解在数学上的存在性、唯一性和稳定性。至于物理含义，显然，不同的问题有不同的物理意义。而存在性，是指验证所求得的解是否满足方程。唯一性，是指讨论在什么样的定解条件下，对于哪一函数类，方程的解是唯一的。通过对唯一性问题的研究，可以明确，对于一定的方程，需要多少个以及哪些定解条件才能唯一地确定一个解。稳定性，是指讨论当定解条件有微小改变时，解是否也只有微小变化。若是，解就是稳定的。对于这个问题的讨论尤为重要，因为在把一个物理问题表达成数学问题时，一般作了一些简化或者理想化的假定，与真实情况有出入。研究了稳定性问题，就可以对解的近似程度做出估计。

一个定解问题，若其解是存在、唯一且是稳定的，就称为是适定的，即在物理上是适当而确定的。

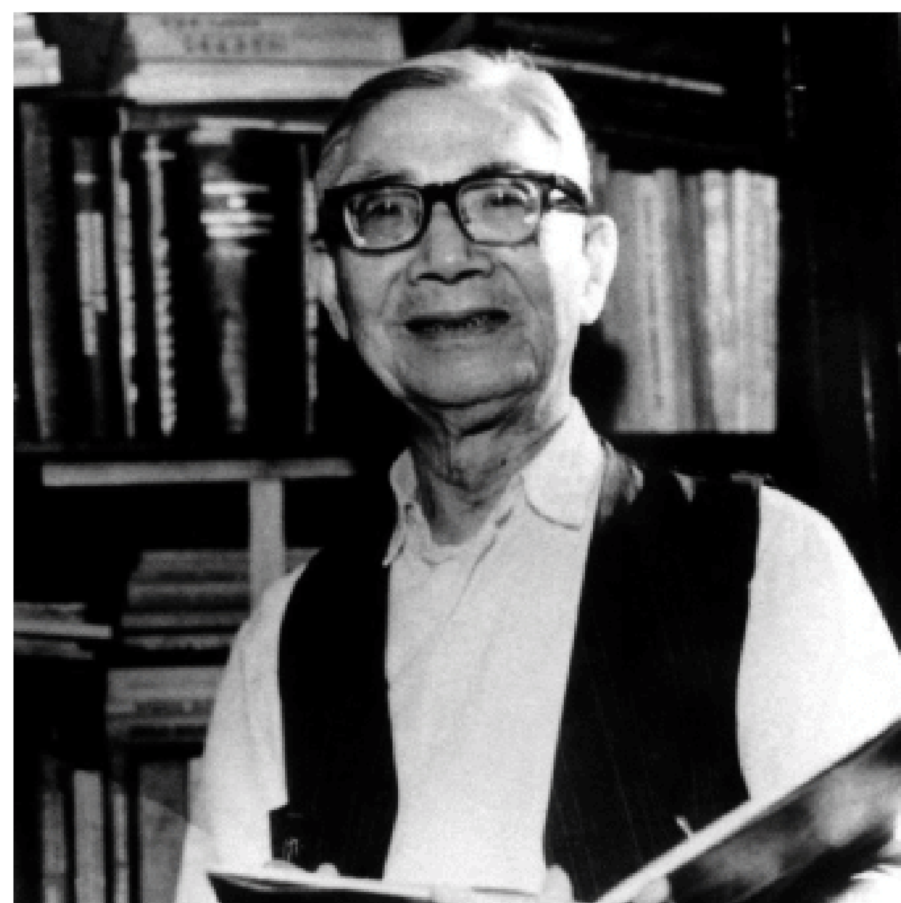
**3. 数理方程的特点**

数理方程一方面紧密地、直接地联系着物理学中的许多问题，另一方面它又要广泛地运用数学中许多内容，所以，它成为数学理论与物理学的实际问题之间的桥梁。在应用数理方法时，要密切注意它的内容与物理学中各种有关现象的联系；要能熟练地掌握所讲述的解决问题的方法，并根据问题的性质来确定使用的方法；要灵活地把各数学分支的知识运用到数理方程中来。因而，这门学科对物理和数学的要求较高。

## 学习资源

### 一、中国近现代著名科学家(物理学)简介

茅以升(1896—1989)，字唐臣，江苏镇江人，土木工程学家、桥梁专家、工程教育家，中国科学院院士，美国工程院院士，中央研究院院士。中国近代桥梁事业的先驱、铁道科技事业的开拓者，同时也是中国土力学的开拓者、科普工作者。



作为我国多学科卓越专家和现代桥梁工程先驱，20世纪30年代，他在钱塘江上建成了中国人自己设计和施工的第一座现代钢铁大桥——钱塘江大桥，在中国桥梁工程史上树立了一座不朽的丰碑。茅以升在钱塘江大桥工程技术上的成就曾令国外同行对中国的桥梁建筑工程师刮目相看。50年代，在武汉长江大桥建设过程中，茅以升担任由中外专家组成的技术顾问委员会的主任委员，解决了武汉长江大桥建设中的14个难题。他还参加了新中国第一座现代化大桥——武汉长江大桥的建造。1959年担任人民大会堂结构审查组组长，为这个历史性的重大建筑贡献了自己的技术、经验和智慧。茅以升是中国现代桥梁工程学的重要奠基人。作为一名教育家，茅以升在教育界工作的二十余年中当过五所学校的教授、两个大学的校长、两个学院的院长。他积极倡导科普教育，撰写了《桥话》《中国石拱桥》《桥梁次应力》《钱塘江桥》《中国的古桥与新桥》等大量的科普文章。1979年应邀访问母校卡内基·梅隆大学时，校长授予他“卓越校友”奖章，以表彰他对世界工程技术方面做出的贡献。1982年被美国工程院授予外籍院士称号。

## 二、诺贝尔物理学奖简介

**1901 年诺贝尔物理学奖**。X 射线的发现。获奖者：W. C. 伦琴（德国）。评价：伦琴于 1895 年在研究阴极射线时发现了这种射线，不过终其一生他也没有弄明白其中的原理，当然他也不需懊恼难过，这超越了他的时代，而这次发现似乎在启示人们，大自然还有更多的 X 因素，我们才只知道了点皮毛而已。

**1902 年诺贝尔物理学奖**。塞曼效应的发现和研究。获奖者：H. A. 洛伦兹、P. 塞曼（荷兰）。评价：所谓塞曼效应是说原子光谱线的光谱线分裂，塞曼于 1896 年观察到此现象后由他的导师洛伦兹用传统的电磁理论解释此现象（洛伦兹终其一生都在修补经典物理，相对论里的公式也是洛伦兹提出来的，但他只求如何迁就传统而没有爱因斯坦般打破一切的勇气，当然这并不表示他不伟大），因此两人共同获得了 1902 年的物理学奖。不过这也是传统理论最后的威风了，随着人们研究的深入发现反常塞曼效应，电磁理论已经无能为力，必须求助量子力学。而洛伦兹这个经典理论的修补匠也难扶大厦于不倒。

应用价值：计算荷质比，测量电子磁场等。所谓光谱，就是某种原子的电子被激发后吸收或释放的固定某几种波长的光线，是当时一种物质的一个基本特征，很多元素的发现就是通过分析光谱得到的。至于物理解释需要通过能级的理论才能很好地证明。

**1903 年诺贝尔物理学奖**。放射性的发现和研究。获奖者：A. H. 贝克勒尔（法国）、P. 居里、M. 居里（法国）。评价：现在大家可谓谈核色变，不过对于认识这个世界来说却是一大进步，当伦琴的 X 射线的论文和照片被贝克勒尔看到后，他深受鼓舞，猜想 X 射线的机制是否和可见光一样，于是就致力于荧光和磷光的实验。不过无心插柳柳成荫，最终发现了穿透力比 X 射线更强的放射性粒子（$\alpha, \beta, \gamma$）。而且通过放射性反应发现原子也是可以转化的，这打破了原子是最小的反应单位的传统理解，而居里夫人发现青铀矿的放射性比纯粹的氧化铀强四倍多，猜测其中有另一种放射性物质，由此开始了艰苦的提炼沥青的历程。这个故事想必大家并不陌生。不过科学的道路并不平坦并且充满危险，就像当年做风筝试验而献身的科学家一样。由于早期没有意识到放射性物质的巨大损害，贝克勒尔健康受到严重损害，于五十多岁离开了人世。

**1904 年诺贝尔物理学奖**。从事气体密度的研究并发现氩元素。获奖者：J. W. 瑞利（英国）。评价：随着科学不断向前发展人们已很难像牛顿那样在物理学的各个方面取得成就。而瑞利恐怕是 20 世纪为数不多的几个全能科学家之一。他的研究几乎包括物理学的各个领域，其中一朵乌云——紫外灾难就和瑞利有关。此外在光学上，他也有巨大成就，关于衍射的瑞利判据可以说是制作照相机、望远镜等光学助视仪器分辨本领的基础。而他发现氩元素的经历可以说几乎被各种教科书反复引用，由于他抓住了不同方法测量氮气密度那极小的完全可以被忽略的误差（五千分之一），导致了惰性气体的发现，他也在诠释他的名言：一切科学上的最伟大的发现，几乎都来自精确的量度。

**1905 年诺贝尔物理学奖**。阴极射线（高速电子流）的研究。获奖者：勒纳德（德国）。评价：勒纳德的获奖让人明白跟好一个老板多么重要，勒纳德曾经给物理学家赫兹作助手，而赫兹让他看了自己的一项新发现：将一块被铝箔包着的含铀玻璃片放入电管中，当阴极射线轰击这块铝箔时，铝箔下面发出了光。当时赫兹以为可以用一片铝箔将空间隔开，一边是按普通方法产生的阴极射线，而在另一边则是纯粹状态下的阴极射线。这个实验以前从未做过。赫兹太忙了，没有时间做这个实验，就让勒纳德做，勒纳德因此在一条指明的路上发明了勒纳德窗的装置，并通过这个装置对阴极射线做出了许多重要研究。相对于他的物

理成就,他的品德一直被人怀疑。他爱慕荣耀,一直觉得自己怀才不遇(尽管他获得了很多荣誉),他还是希特勒的物理学顾问。

**1906年诺贝尔物理学奖**。气体导电。获奖者:J.J.汤姆孙(英国)。评价:所谓气体导电产生的就是阴极射线,关于阴极射线的本质争论已久,而汤姆孙用实验证明了阴极射线就是带电粒子,而且计算出了阴极射线的质荷比,并将这种粒子命名为电子。他还建立了最早的原子模型,蛋糕加葡萄干模型。汤姆孙少年得志,年仅27岁就担任著名的卡文迪许实验室主任并培养了一大批伟大的物理学家。当然他也培养了他的儿子,成就了父子皆获诺贝尔奖的佳话(卡文迪许实验室建于1871—1874年间,是当时剑桥大学的一位校长威廉·卡文迪许私人捐款兴建的。由麦克斯韦负责建立,培养了大批物理学家:瑞利,汤姆孙,卢瑟福等)。

**1907年诺贝尔物理学奖**。光学精密计量和光谱学研究。获奖者:迈克耳孙(美国)。评价:关于物理学的两朵乌云有一个就是以太的存在性,而迈克耳孙就是最成功地证明以太不存在的人(虽然他也不想)。所谓以太论就是说空间中存在一种物质以太,是一种绝对静止的参考系,光就是通过这种介质传播的。不过迈克耳孙和莫雷做了著名的实验否定了以太的存在,因为实验证明光朝各种方向相对于地球的传播速度相等。

应用:试想要测量光相对于地球公转的速度,那是何等悬殊的数值,必须有精密度极高的器材才行。因此迈克耳孙干涉仪利用光波波长作为测量单位也是测量上的革命。把人类的测量精度提高了两个数量级。也为后来的光学测量打下了基础。

**1908年诺贝尔物理学奖**。照片彩色重现。获奖者:G.李普曼(法国)。评价:关于照相技术的革新曾经两次获得诺贝尔奖,李普曼通过利用不同波长的干涉图样的不同,记录了彩色(其实不同颜色的光就是波长不同而已)信息。虽然这种方法现在被淘汰,但是李普曼让我们看到了彩色的照片。其实光的干涉不仅可以记录彩色信息还可记录立体信息,这点会在日后的全息照相中提及。光的干涉首先由托马斯·杨发现:干涉是两个同频率,有相同振动方向(或分量),存在固定位相差(即固定的距离)的波互相加强或减弱的现象,和衍射一样是波的基本属性。

**1909年诺贝尔物理学奖**。无线电报的发明。获奖者:G.马可尼(意大利)、K.F.布劳恩(德国)。评价:无线电报的发明有很强的实用色彩,无线电报就是通过无线电磁波作载体来通信。在今日的社会可谓无人不知无人不晓。毫无疑问无线通信是人类历史的革命,使世界大大变小了。这两人都是在应用物理学上有突出成就的人,马可尼发明并不断改良了无线电报系统的传播距离并最终实现了越洋电报。布劳恩通过坚强的物理基础对通信系统装置进行了改进。两人相比较,可以说马可尼像个发明家,而布劳恩是一位科学家。

**1910年诺贝尔物理学奖**。气液状态方程。获奖者:范德瓦耳斯(荷兰)。评价:气态方程 $pV=nRT$ 是我们中学常用的一个公式。可惜这是针对理想气体而言,与实际情况还是有很大的出入。而范德瓦耳斯提出了新的方程:$(p+a/V^2)(V-b)=RT$;其中,$p$、$V$ 和 $T$ 分别代表气体的压强、体积和温度,$R$ 是气体常数,$a$ 代表分子之间的吸引的唯象参数,$b$ 为分子的体积,且 $a$,$b$ 对于不同的气体有不同的值。根据这个方程计算的结果与实验数据符合得很好。为了纪念他把分子间力称为范德瓦耳斯力。

## 思考与练习

1. 谈谈物理实验方法在物理学发展中的巨大作用。
2. 理想化方法的分类。

3．比较与类比方法的区别与联系。
4．举例说明模拟方法在物理学研究中的重要作用。
5．谈谈归纳与演绎的特点。
6．谈谈分析与综合的特点。
7．物理学有很多“假说”存在了很长时间了，为什么一直称为假说？
8．为什么说除了实验方法之外，数学方法是物理学理论发展、建立和完善的根本方法？

# 第3章

# 物理学中的七个基本量

计量学在18世纪以后作为一门学科出现。当伟大的科学家牛顿创立牛顿力学使经验科学上升为理论科学之后，计量学才成为一门专门学科而出现在科学技术领域。从前，人们对测量本质还未形成清晰的认识，不明确客观事物之间的联系，因此创造了大大小小互不关联的计量单位和各种各样的测量结果的表示方法。这些成千上万的单位给人类的工作和生活带来极大的不便。随着对客观事物认识的逐步深入，人们发现各种物理量之间普遍地存在着一定的联系。因此，单位制的概念也就逐渐地形成了。

物质世界的存在状态和运动形式是多种多样的，自然科学的任务在于探索物质存在和运动的规律。量就是阐述物质存在和运动规律的一个最重要的基本概念。量的具体意义是指大小、轻重、长短等概念，而量的广义是指现象、物体或物质的定性区别，即根据量可以区别长度、质量、时间、温度等物理量。量是由数值和计量单位的组合来表示，用没有计量单位的纯数来表示量的大小是毫无意义的。换言之，量的表示必须在其数值之后注明所用的计量单位。由一个数和计量单位的乘积表示的量称为量值。

测量是人们揭示自然规律的手段，借以定量描述和定性区别周围物质世界，从而达到改造客观世界的目的。人们在广泛的社会活动中，每日每时都在进行着各种不同的测量。例如，做衣服要量布，生病要量体温。在生产和科技领域，测量就更广泛，一切数据无不经测量取得。最初的测量方法是原始的，单位也是可以任意选定的。随着生产的发展和商品交换成为社会性活动，客观上就要求测量的统一，即对同一物体，在不同的地点，用不同的方法进行测量，要求其测量结果在一定的准确度内是一致的。随着科学技术的飞速发展，计量制度的混乱严重地妨碍着生产、科研、文教、经济事业的发展。为此就要求以法定的形式建立统一的计量标准，从而形成了有别于测量而称之为“计量”的新概念。在现代计量学中，定义“测量”为以确定被测对象量值为目的的全部操作；定义“计量”为实现单位统一和量值准确可靠的测量。

一个客观量总是由数值和计量单位组成的。对某一量进行测量，首先要选定计量单位，即选定与待测量同种的特定量，测量过程就是将待测量和该特定量进行比较的过程。计量单位是可以选择的。如果选定的单位不同，同一物理量的数值大小也不同。由于这些单位间存在严格的换算关系，使用不同单位测量同一物理量的数值大小尽管不同，但它们所表示的物理量的量值大小是一致的。物理量的种类很多。如果对每一个物理量都单独选定一个单位，并建立相应的标准，那是很不经济的，同时也容易引起这些物理量之间关系的混乱。人们早就认识到，许多物理量之间存在着严密的逻辑关系。因此，可以选定几个彼此独立的物理量单位为基础，作为基本单位。从这些基本单位出发，通过物理量之间严格的相互关系（包括物理量的定义公式和物理规律的表述公式），就可以推导出所有其他物理量单位，这些单位称为导出单位。由基本单位和导出单位构成的物理量单位体系叫作单位制。

在 1971 年第十四届国际度量衡大会(General Conference of Weights & Measures,CGPM)中,选择了 7 个物理量作为基本量的国际单位系统,其法文名称"Le Systeme International d'unites"缩写为"SI",我国简称其为国际制。

本章简要介绍国际制中的 7 个物理量历史演变、定义和测量方法。

## 3.1　长度

### 3.1.1　米制前的长度单位

在人类最早的文明中,由于生活和生产的需要,很早就开始建立计数和表示数目的各种计量制度。在不同时间、不同地区所形成的计量制度是形形色色的,甚至同一单位表示的量的大小都可能相差很大。15、16 世纪欧洲资本主义生产关系的形成和近代工业的发展,推动了贸易的发展和科学技术的进步。系统的观察和严密的数学演绎相结合的研究方法被引进到物理学研究中,导致了 17 世纪主要在力学和天文学领域的科学革命,建立了牛顿力学体系。到 18 世纪,力学、光学、热学和静电学均已完成奠基性的工作,成为物理学的几门基础科学。人们的测量范围也从简单的长度、容量和质量扩展到所有的力学量、电学量和光学量等。科学家们开始寻求一种更适合的国际通用单位,并以它为基础得到一种所有国家都认可的计量制度。为了建立统一而合理的计量制度,需要考虑以下三个重要方面:一是作为基本单位定义的自然物应该是永恒而不变的;二是制作或复现标准器的材料应有尽可能好的时间稳定性;三是同一种类的各单位之间的进位应简单。对于长度的基本单位,当时有两种完全不同的建议,其一是由法国天文学家莫顿(William Morton,1819—1868)和惠更斯提出的采用秒摆摆长作为长度单位的建议。但由于秒摆的摆长与重力加速度有关,因此,不同地点的秒摆摆长是不同的。当时法国科学院赞同另一个建议,即采用与地球子午线长度相联系的一种基础长度作基本单位。经法国科学院的推荐,法国国民大会于 1791 年采纳了只基于一个长度单位"米"的计量制度原理。"米"被定义为等于地球子午线四分之一长度的一千万分之一。面积和体积的单位分别是平方米和立方米的十进倍数单位和分数单位。质量(重量)的单位是 1 立方米($m^3$)的水在密度最大时(温度为 4 摄氏度(℃)时最大)的质量(重量)。由于这一制度以米为基础,故得名"米制"。

法国人开创米制后,米制单位即开始向欧洲和世界传播。长度单位"米"是国际单位制中最重要的基本单位之一,也是米制单位的基础。自 1889 年第一届国际计量大会正式确定米的长度以来,对米共作过三次规定。目前采用的是历史上的第三个米定义,是 1983 年第十七届国际计量大会正式通过的。

**1. 历史上各种长度单位**

由于世界各地文化、历史和生产力发展等方面的差别,各国早期采用的长度单位也大相径庭。古代各国大都采用人体某一部分的长度作为计量单位。例如,古埃及人将肘关节到指尖的距离定为长度单位,称为"腕尺",其长度约为现在的半米。九世纪的英国,规定以国王亨利一世的鼻尖到手臂平伸后指尖的距离定为 1 码。英尺是英王查理曼大帝的脚长,而英寸则是英王埃德加的大拇指关节长度。法国的尺称为"脱瓦斯",是按国王脚掌的六倍长度确定的。

据《大戴礼记·主言篇》记载,我国古代有"布手知尺"之说,即以拇指和食指张开的距离定为 1 尺。从"尺"字的写法,很容易看出该字是根据手在测量物体长度时的形象创造并转

化而来的(图 3-1-1)。

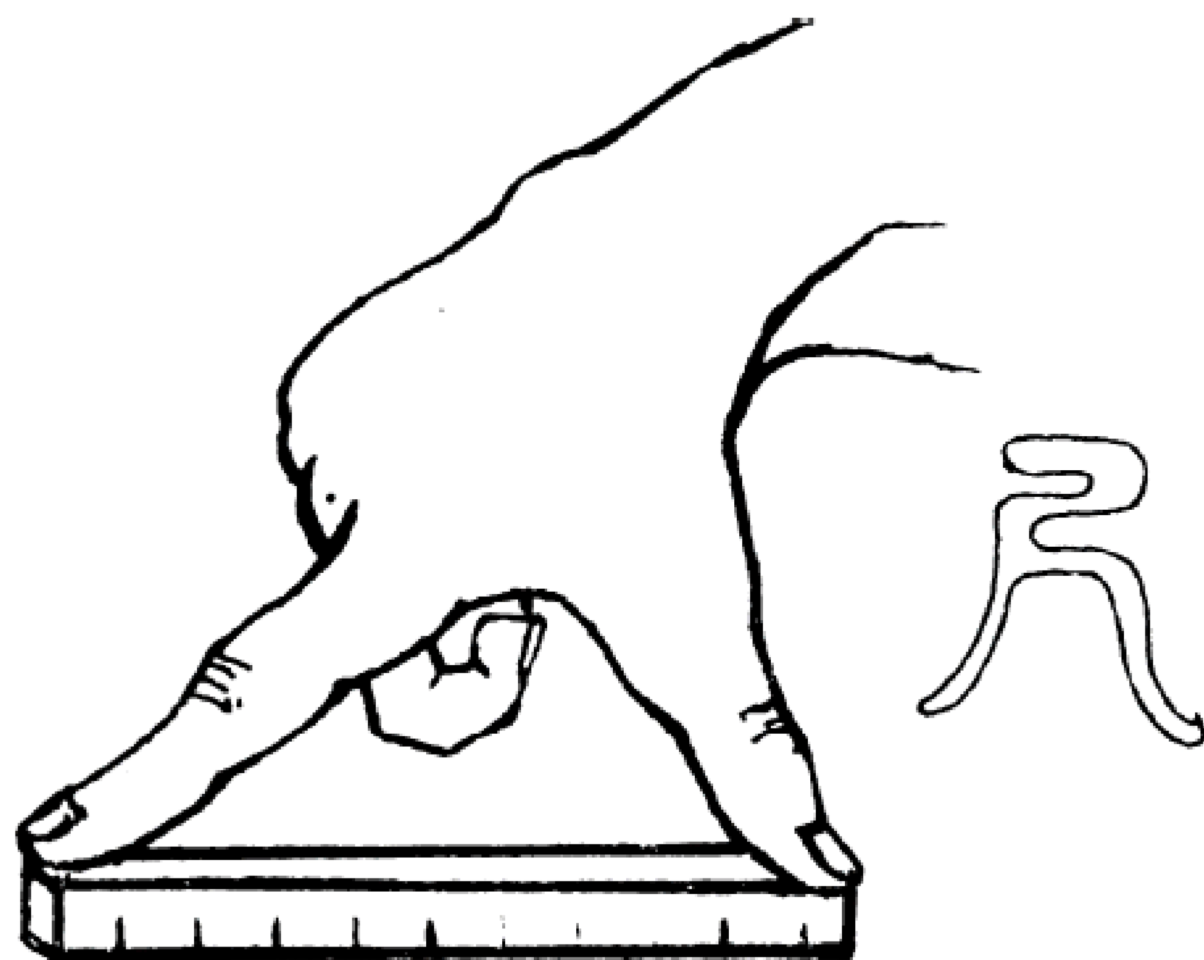


图 3-1-1 布手知尺

由于人体的身高和各部分器官的大小因人而异,且相差甚大,此后就慢慢发展到用较稳定的自然物来代替。我国汉代规定"以子谷秬黍中者,一粒宽为一分,十分为一寸,一百分为一尺"。据 1305 年英国《法令汇编》,英尺又定义为"自穗之中部,取大麦三十六粒,头尾相接排列之长度",两者非常相似。16 世纪的德国,曾将长度单位定为"星期日立于教堂门首,礼拜完毕后,令最先走出教堂之男子 16 名,高矮不拘,随遇而定,各出左足前后相接,取其长度十六分之一"。

我国古代某地区还有一种长度标准,称为"黄钟律管"。黄钟为我国古代乐理中十二律之首,大体相当于现代乐谱中 C 调的"1"。它是一支厚薄均匀的竹管,其管长和内径一定。若将声波传播速度视为不变,则声波波长与频率成反比。在内径一定的情况下,律管越长,声波波长也就越长,而频率(即音调)也就越低。这样,凡是能吹出黄钟音调的律管,其长度基本一致,所以,可将它作为长度标准。这完全符合近代物理原理,可以说这是世界上最早的自然基准。

**2. 我国历史上长度单位的变化**

我国历史上,战国时期诸侯割据,各国的度量衡单位名称和进位制度都不统一。秦始皇统一六国以后,以当时的秦制统一了全国的度量衡。但由于种种原因,在各个不同的历史朝代,仍有着各自不同的计量单位。我国历代的长度单位虽然都称为"尺",但它们实际的长度的变化仍是相当大的。

由中国历代尺长值表 3-1-1 可知,商朝尺最短,每尺约合 16 厘米(cm),而民国时代每尺为 33.33cm,两者相差竟达 17cm 之多。一般来说,古代的尺要比近代短不少,无怪乎在一些古典小说中经常会有"身高八尺"之类的描述。

**表 3-1-1 我国历代尺长的变化** cm

| 时代 | 年代 | 尺长 | 时代 | 年代 | 尺长 |
|---|---|---|---|---|---|
| 商 | 公元前 16 世纪到公元前 11 世纪 | 16 | 南北朝 | 420—589 | 南朝 24.7<br>北魏 28<br>北周 29<br>东魏 30.2 |
| 战国 | 公元前 475—前 221 | 23.1 | | | |
| 秦 | 公元前 221—前 207 | 23.1 | | | |
| 西汉 | 公元前 206—8 | 23.1 | 隋 | 581—618 | 29.5 |
| 新 | 9—25 | 23.1 | 唐 | 618—907 | 30.3 |
| 东汉 | 25—220 | 23.5 | 宋(辽、金) | 960—1279 | 31.6 |
| 三国 | 220—265 | 24 | 元 | 1271—1368 | 31.6 |
| 两晋 | 265—420 | 24.4 | 明 | 1368—1644 | 32 |
| | | | 清 | 1644—1911 | 32 |
| | | | 中华民国 | 1911—1949 | 33.33 |

**3. 米的起源**

各种单位和计量制度的混乱,对科学和贸易的发展起了阻碍作用。据 1742 年的测量结

果,法国的长度单位就比英国长百分之六。从 17 世纪开始,科学家们就试图建立一种能为大家所接受的国际通用计量制度。此外,由于实物基准容易受到诸如战争、火灾等偶然事件的破坏,人们也一直希望能采用自然界的不变长度作为单位。也就是说,希望采用自然基准而不是实物基准。英国最初的码原器就被毁于 1834 年英国议院的一次火灾,此后不得不重新制造其码原器。

法国在确定采用米制后,历时六年,测量了从法国敦克尔刻到西班牙巴塞罗那间巴黎的子午线弧长。根据这一测量结果,制造了一根铂制端面标准尺,用以体现米的长度。由于该尺保存在法国档案局内,故也称为“档案局米尺”。

事实上,不久就发现,档案局米尺偏离了原来的定义,它比地球子午线长度的四千万分之一短了约 0.2mm。但因为地球子午线长度不易测量,因此就直接用档案局米尺作为米的标准,并在法国使用了将近 100 年。

## 3.1.2 米的定义

由于米制具有简易、逻辑结构合理、通用性广等有优点而逐步为其他国家所接受。由法国政府主持,在 1869 年和 1872 年两次召开“国际米制委员会”,会议决定制造新的国际基准米尺,即米原器及其复制品。新原器的长度以档案局米尺为基准,其复制品则分发给米制公约各参加国,作为各国的基准米尺。新的米原器用含铂 90%、铱 10%的合金制造,其材料的稳定性完全可以满足米原器的要求。其“X”形的截面无疑是最坚固和用料最省的形状。即使米原器有些弯曲变形,对尺长的影响也将很小。由于增大了尺子的表面积,较容易达到与周围环境的温度平衡。将端面基准改为线纹基准,可以防止因磨损而引起的尺长变化,同时线纹尺之间的对比也较端面尺容易进行。国际米原器的横截面的外接正方形每边长 200mm,“X”形尺子(见图 3-1-2)的中间平面宽 40mm,在中间平面的两端各有 8mm 长的抛光部分,其上刻有间距为 0.5mm 的三条刻线,中间刻线距端面 10mm。当米原器处于冰点温度时,两条中间刻线的轴线之间的距离定为米。两条中间刻线的有效部分是由两条间距为 0.2mm 的轴向基准划定的。国际米制委员会制造了 31 根同样的米原器,其长度都与档案局米尺作了比较,其中六号米尺的长度和档案局米尺最接近,因此就将它作为国际基准米尺,并保存在国际计量局。

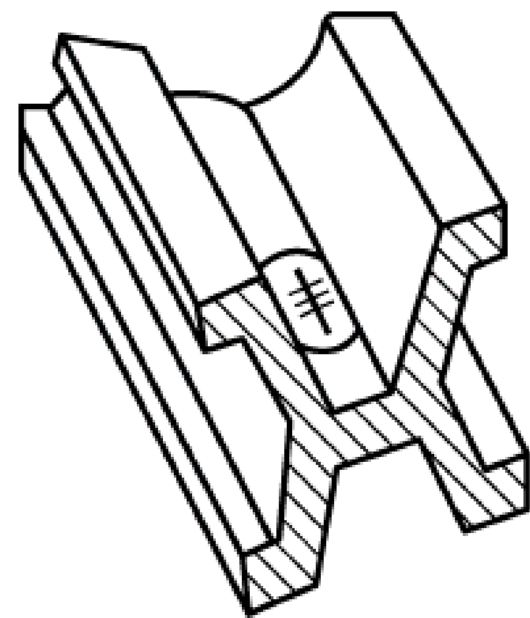


图 3-1-2 国际米原器

**1. 第一个米的定义——光波波长标准**

1889 年第一届国际计量大会批准了国际计量委员会所选择的米原器,并宣布“该米原器在冰融点温度时代表长度的米单位”。考虑到环境因素对尺长的影响,1927 年第七届国际计量大会又对此作了更明确的规定:“长度的单位是米,规定为国际计量局所保存的铂铱尺上所刻的两条中间刻线的轴线在 0℃时的距离,这根铂铱尺已被国际计量大会宣布为米原器,保存在标准大气压下,放在两个对称地置于同一水平面上并相距 571mm 的直径至少为 1cm 的圆柱上。”

该定义规定了米原器的工作温度,消除了热膨胀对测量结果的影响;规定了应在标准大气压下使用,消除了由于气压变化导致了尺长的变化;规定了米原器放置时两个支点的距离,使由尺子自重引起弯曲而使尺长变短的程度为最小。

**2. 用$^{86}$Kr 橙色谱线定义米**

当时对$^{114}$Cd、$^{198}$Hg、$^{84}$Kr 和$^{86}$Kr 等核素的光谱进行了仔细的研究,并测量了它们的谱

线宽度,最后选择了宽度最窄的$^{86}$Kr 橙色。其谱线宽度为 $5\times10^{-4}$ 纳米(nm),波长不确定度为 $1\times10^{-8}$。它比米原器或镉红线的准确度高约一个数量级,因此,在 1960 年第十一届国际计量大会上正式批准废除铂铱米原器,而将米定义改为:“米等于$^{86}$Kr 原子的 $2p_{10}$ 和 $5d_5$ 能级间的跃迁所对应的辐射在真空中 1650763.73 个波长的长度。”

根据这一定义,$^{86}$Kr 基准辐射的真空波长值应是 605.780211nm。这是通过与镉红线进行波长比对而确定的。由于波长比对是在真空中进行的,故应先将镉红线的空气波长换算到真空波长。为此,国际计量局推荐了通过气压、温度和空气湿度计算空气折射率的经验公式。

为了使这一定义更完善,国际计量局还通过提议,规定了产生$^{86}$Kr 基准辐射的放电管形式、参数和使用条件,并认为在满足使用条件时,其辐射波长在不确定度为 $1\times10^{-8}$ 范围内等于未受扰动原子所发射的波长。

**3. 用光速定义米**

米定义咨询委员会提出了重新定义米的下述理由,表明更改米定义的时机已经成熟。

(1) $^{86}$Kr 米定义已经不能足够精确地复现米的长度以满足各方面的需要;

(2) 激光稳频技术的进展,使激光的复现性和易于应用方面已大大优于$^{86}$Kr 基准;

(3) 由于激光频率和波长测量的进展,可以更精确地确定光速,但其准确度却受到$^{86}$Kr 米定义的限制;

(4) 由激光频率测量及给定的光速值所导出的激光波长的准确度比$^{86}$Kr 基准辐射更好;

(5) 在天文和大地测量领域,保持光速值不变具有重要意义,这样用光速测距而得到的距离数值不必随光速值的日益精确而不断修正;

(6) 只要新的米定义隐含光速值等于目前的推荐值,则新定义的长度单位和原有定义将保持连续性,即两者之差别小于$^{86}$Kr 基准的复现准确度 $4\times10^{-9}$;

(7) 已对新米定义的各种叙述形式进行了深入的讨论,认为这些形式本质上是相同的,并已取得基本一致的意见,即用光在某一特定的时间间隔内所通过的行程来定义;

(8) 目前已经能够给出一个比较实用的米定义表达方法,包括如何对待现在作为长度基准使用的$^{86}$Kr 辐射。

以上充分清楚地阐明了用光速定义米的可能性、必要性及新米定义的形式等问题。

1983 年第十七届国际计量大会通过了新的米定义,其表达如下:“米等于光在真空中 299792458 分之一秒时间间隔内所经路径的长度。”

现就该定义的某些物理含义作如下讨论:

(1) 定义隐含了光速值 $c=299792458\text{m/s}$,这是一个定义值,因此,是一个精确的、不具有误差的数值。

(2) 如果已知激光频率,则激光波长可以由频率和光速值导出,并具有和频率测量相同的不确定度。这样做的好处是频率测量往往具有比波长测量更高的准确度。

(3) 定义中用了“光”,而不采用“平面电磁波”的说法。按说,平面电磁波似乎更具有普遍性。采用“光”一词的主要原因是有些科学家担心光速是否也存在色散,即光速可能与频率有关。当然,即使光速有色散效应,估计也是十分小的,目前的技术水平还不可能发现这种差别。为了保持定义的严格性,并使它在较长时间内不需要修改,采用“光”的说法比较合适一些,因为它仅指某一频段内的电磁波。

(4) “真空”一词指不存在任何实物粒子。曾有人提议采用“自由空间”的说法,其含义

是指无场空间。严格地说，光的传播会受到场的影响。例如引力场，因而采用自由空间的说法似乎更确切一些。不采用自由空间一词的原因是无场空间很难实现，而且目前在理论上对于零场的修正也没有完全解决。比较而言，"真空"一词更容易被理解和接受。

**4. 长度量值的复现方式**

第一次米的定义是用铂铱合金米尺来复现的，第二次米的定义是用$^{86}$Kr 基准谱线的波长来复现的，新的米的定义，根据国际米定义咨询委员会的推荐，可用下列三种方法复现：

(1) 若平面电磁波在时间间隔 $t$ 内，在真空中所经路径的长度为 $l$，则根据测得的时间间隔 $t$ 就可以得到 $l$，其关系式为：$l=ct$。其中真空中的光速为 $c=299792458\text{m/s}$ 是直接从定义中得到的无误差的精确值。

(2) 测量出某一平面电磁波的频率 $\nu$，则可以由光速值得到其波长值，$\lambda=c/\nu$。

(3) 采用推荐的辐射表中的任何一种辐射，表中给出了它们的频率值、真空波长值，以及相应的不确定度。同时，还对产生这些辐射的有关参数和使用条件作了规定。

### 3.1.3　长度测量应用举例

**1. 激光测距仪**

用激光测距仪可以进行几十米到几千米的大距离测量。其原理类似于拍频波干涉仪，但其对应于标准波长的频率不是由双纵模的频差得到，而是直接对激光辐射进行高频调制而得到的。激光将该调制波发射出去，通过测量回波的位相就可以确定干涉级次的小数。其整数级次则采用小数重合法确定。改变调制频率，即相当于采用不同的波长进行测量，最后根据测到的一组小数就可以确定被测距离。

**2. 光速测距：地球—月球距离的测量**

地球和月球相距 384000km，故十分适于直接利用光速测距。即从地球上发射一束光，测量出从发射到反射回来的时间间隔，就可以由光速算出月球与地球表面的距离。

在激光出现以前，用这种方法测量地-月距离是不可能的。设想有一直径 1cm 的光源，并置于焦距为 1cm 的抛物面镜焦点处而形成探照式的平行光束。当光束到达月球时，其直径扩展到 3800km，比月球直径 3476km 还要大。这时月球表面的光束能量密度降低到抛物面镜处的 $10^{-13}$。如果在月球上也设置一直径为 1m 的平面镜，将来自地球的光再反射回地球，则能量密度将再次减少 $10^{-13}$。这就是说，光在地球和月球间往返一次后，其光强将减弱到原先的 $10^{-26}$。显然，检测如此微弱的光是十分困难的。

如果利用激光，则情况就不同了。激光具有很好的方向性，对于直径 1m 的 633nm 平行光束，由衍射效应引起的发散角仅为 0.2°，到达月球时直径为 300m。同样用直径 1m 的平面反射镜使激光束返回地面，则地球上接收到的光强将比用探照灯提高 $10^{16}$ 倍。

精密测量地—月距离，要求确定月球上被测点的位置，同时还应在该处设置反射镜以加强反射回地球的光信号。这一问题被 1969 年美国发射的阿波罗 11 号登月飞船所解决，它将镶在边长 46cm 正方形铝板上的由 100 个直径 38mm 的立体棱镜组成的反射器送到月球表面，作为测量地-月距离的目标，安置在月球表面的不同位置。图 3-1-3

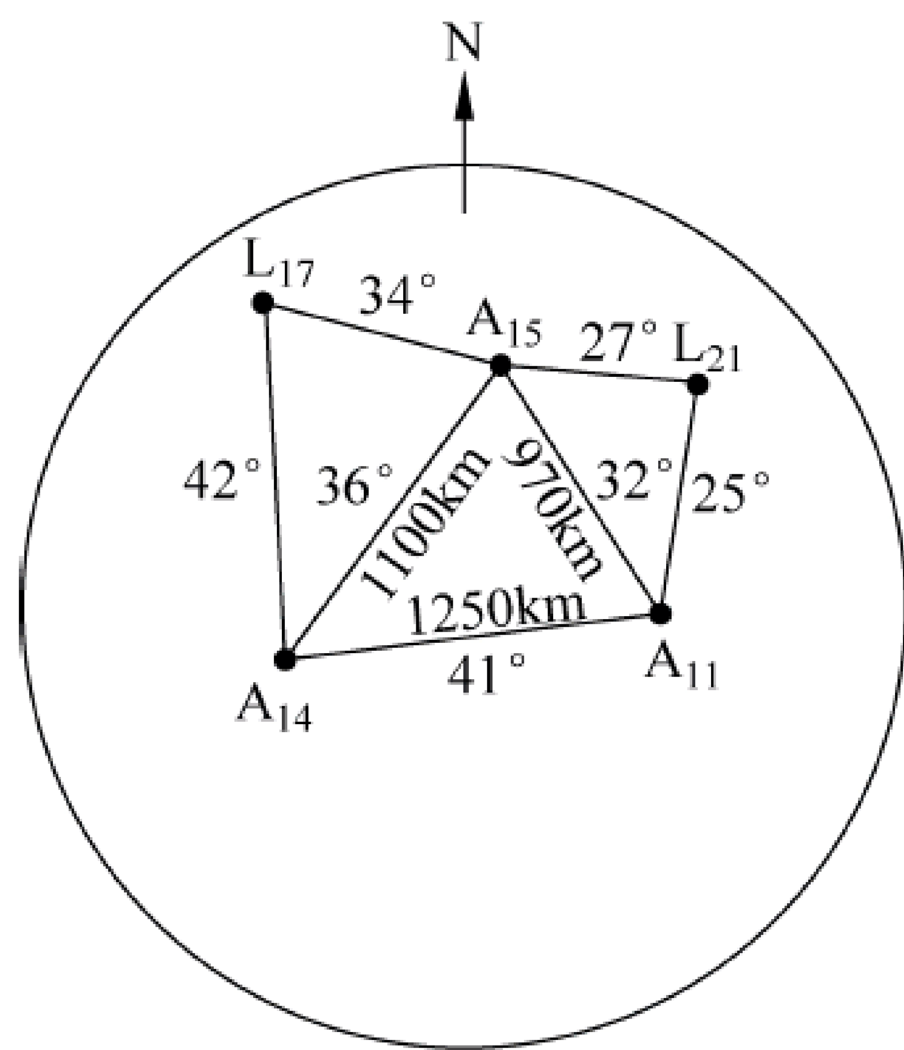



图 3-1-3　放置在月球上的各反射器的位置

给出放置在月球上的各立体棱镜反射器的位置分布,图中示出各反射器之间的角度是月球中心与各反射器相连接的半径夹角,同时也示出了各反射器之间的距离。$A_{11}$ 即是阿波罗11号送去的反射器。

美国麦克唐纳天文台1970年用直径2.7m的反射望远镜将脉冲红宝石激光束射向月球,在约2.5s以后检测到来自月球的反射光信号,并以1ns的准确度对此时间间隔进行测量。测量结果表明各脉冲光的往返时间约有4ns的起伏,这对应于被测距离约有1.2m的不确定度。若取多次测量的统计平均值,其测量不确定度可以缩小到30cm,这相当于$8\times10^{-10}$的相对不确定度。这一测量准确度已可以用来发现由于地球和太阳的引力而引起的月球表面不规则的运动。

## 3.2 质量

质量是物理学最基本、最重要的概念之一。随着人们对经典物理现象及概念逐步深入的研究认识,对物质质量的理解、界定和测量方法经历了一个漫长的发展变化过程。1960年,第十一届国际计量大会通过的国际单位制(SI),将质量确定为7个基本物理量之一:其名称为"质量"(mass),简写为$M$或$m$;其单位名称为"千克",国际单位代号为"kg";并作文字定义:千克等于国际千克原器的质量。

### 3.2.1 国际千克原器

**1. 国际千克原器的由来**

最初的千克质量单位是由18世纪末法国采用的长度单位米推导出来的。1立方分米纯水在最大密度(温度约为4℃)时的质量,就定为1千克。1799年法国在制作铂质米原器的同时,也制成了铂质千克原器,保存在巴黎档案局里。后来发现这个基准并不准确地等于1立方分米最大密度纯水的质量,而是等于1.000028立方分米。于是1875年米制公约会议之后,也用含铂90%、铱10%的合金制成千克原器,一共做了三个,经与巴黎档案局保存的铂质千克原器比对,选定其中之一作为国际千克原器。这个国际千克原器被国际计量局的专家们非常仔细地保存在特殊的地点,用三层玻璃罩罩好,最外一层玻璃罩里抽成半真空,以防空气和杂质进入容器导致质变。随后又复制了四十个铂铱合金圆柱体,经过与国际千克原器比对后,分发给各会员国作为国家基准——国家千克原器。跟米原器一样,千克原器也要进行周期性的检验,以确保质量基准的稳定可靠。1889年第一届国际计量大会批准以这个国际千克原器作为质量标准,沿用至今。

中国"国家千克基准"在1965年由国际计量局检定,并由伦敦的Stanton仪器公司加以调整,严格保存在北京中国计量科学院的质量标准库中。

众所周知,国际单位制(SI)中有7个基本单位。目前,千克在国际单位制(SI)中是唯一仍根据一种人工材料——国际千克原器而充当基准的基本单位。这实在是和科学技术日新月异的发展形势极不相称,因此人们开始研究质量"kg"的新定义。

**2. 人工实物基准的缺陷**

保存像国际千克原器这样的人工实物基准,在实践中存在许多问题,例如:

(1) 国际千克原器有可能被损坏,甚至毁于战火,如果发生这类事件,后果不堪设想。

(2) 没有严格的科学定义,缺乏可靠的确定性。例如,在国际千克原器上有可能积存外来杂质,有的杂质很难发现,也很难清洗干净,这就是说,无法保证国际千克原器精确无误的

可复现性。

(3) 国际千克原器以尚未掌握的规律老化。

(4) 国际千克原器和国家千克原器只能在一个实验室使用,数值传递非常麻烦。

(5) 为了避免磨损和污染,国际千克原器只能尽量少使用,这样就大大限制了它的使用价值。

### 3.2.2 质量的物理意义

**1. 物体的质量是其含物质的多少**

牛顿对质量的最初定义是物体的质量是其含物质的多少,牛顿指出:"物质的量是用它的密度和体积一起来量度的。"这是一种朴素和直觉的表述。鉴于知识的渐进性和可接受性,目前我国初中物理教科书就是采用定义:"物体中含有物质的多少叫质量"(人教版)。而且还指出"质量不随物体的形状、状态、温度和位置的改变而改变(这一说法并不准确),是物体本身的固有属性"。

**2. 质量表示物体的惯性**

质量是描述物质属性的物理量。物体具有保持原有运动状态的属性——惯性。从定性上,质量是物体惯性的量度,这是牛顿第一运动定律表述的第一层含义(牛顿第一运动定律表述的第二层含义是:力的定性定义)。牛顿第二运动定律表述的是力的定量定义。

牛顿第二运动定律是把物体在其速度变化(运动状态变化)时所显示的阻抗能力(惯性)定量化的量度。

即 $\boldsymbol{F}=\frac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t}=m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}+\boldsymbol{v}\frac{\mathrm{d}m}{\mathrm{d}t}$,由于当时无法解释 $\boldsymbol{v}\frac{\mathrm{d}m}{\mathrm{d}t}$ 一项,即认为 $\frac{\mathrm{d}m}{\mathrm{d}t}=0$,所以有 $\boldsymbol{F}=\frac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t}=m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=m\boldsymbol{a}$,式中 $m$ 即为质量,是物体平动惯性大小的量度,也是物体作加速平动运动属性的表现,因而称 $m$ 为惯性质量,我国现行高中物理教材就采用这一表述方式。

**3. 质量是表示物体间引力大小的量度**

质量是物体具有产生引力作用和感受引力作用的属性。牛顿于17世纪发现的万有引力定律就是反映这一属性的规律,其表示式 $F=G\frac{m_1m_2}{r^2}$,式中的 $m_1$ 和 $m_2$ 就是相互作用的两个物体分别具有产生引力作用(产生引力场——牛顿是持超距作用观点的,不认为有场的存在),并能感受引力作用大小的量度,因而称之为引力质量。

**4. 惯性质量与引力质量的关系**

设地球表面某处一物体,如果忽略地球自转效应,它受到地球的万有引力为

$$F_1=G\frac{m_1M}{R^2} \tag{3-2-1}$$

式中 $M$ 和 $R$ 分别为地球的质量和半径,$G$ 为万有引力常数,$m_1$ 为该物体的引力质量;另一方面,设此物体在力 $F_1$ 的作用下产生的加速度为 $a_1$,则

$$F_1=m'_1a_1 \tag{3-2-2}$$

式中 $m'_1$ 为惯性质量。以同样条件设另一物体,则

$$F_2=G\frac{m_2M}{R^2} \tag{3-2-3}$$

$$F_2=m'_2a_2 \tag{3-2-4}$$

由以上四式得：$a_1=\frac{m_1}{m_1'}\frac{GM}{R^2}$，　$a_2=\frac{m_2}{m_2'}\frac{GM}{R^2}$。

实验事实是：地球表面同一处的重力加速度都是相等的，即 $a_1=a_2=\cdots=g$，因此有 $\frac{m_1}{m_1'}=\frac{m_2}{m_2'}=\frac{m_3}{m_3'}=\cdots$。

由此可见，一切物体自由下落时的加速度都相等的事实，意味着引力质量与惯性质量存在正比关系，在万有引力定律和牛顿第二定律中，如果选取适当的单位，则二者是相等的。正是基于这样的客观事实，爱因斯坦提出了引力场的概念，并认为加速场可以抵消引力场效应，总结出了“等效原理”，作为广义相对论的实验基础并将引力质量和惯性质量用如下关系联系起来：

惯性质量×加速度(加速场强度)＝引力质量×引力场强度

不过，等效应原理的动力学机制至今还不太清楚。

**5. 质量随速度的增长而增加**

其实，在牛顿第二定律 $\boldsymbol{F}=\frac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t}$ 中已经得出了 $\boldsymbol{v}\frac{\mathrm{d}m}{\mathrm{d}t}$ 等于什么的问题，只是牛顿力学是建立在超距作用的绝对时空观上，将质量视为一个与运动无关的恒定量。而在狭义相对论力学中运动物体的质量与其速度有如下关系：

$$m=\frac{m_0}{\sqrt{1-v^2/c^2}} \tag{3-2-5}$$

式中 $m_0$ 为静止质量，即 $v=0$ 时 $m=m_0$，$c$ 为光速。物体高速的运动效应，引起物体质量的改变，说明物体的惯性或产生和感受引力的性质发生了变化，同时也说明了物质的属性与其运动有关。在低速下，由于物质属性的改变极为微小，从而才认为物体的质量是恒定的。

**6. 质能关系是近代物理学的基石之一**

在相对论力学创立之前，人们一直以为质量守恒定律和能量守恒定律是两条彼此没有关系的完全独立的定律。相对论力学揭示了质量与能量之间存在着十分简单的关系。在 $m=m_0/\sqrt{1-v^2/c^2}$ 中，当 $\frac{v}{c}\ll 1$ 时，将其进行幂级数展开得

$$m=\frac{m_0}{\sqrt{1-v^2/c^2}}=m_0\left(1-\frac{v^2}{c^2}\right)^{-\frac{1}{2}}=m_0\left(1+\frac{1}{2}\frac{v^2}{c^2}+\cdots\right)$$

即

$$mc^2=m_0c^2+\frac{1}{2}m_0v^2+\cdots \tag{3-2-6}$$

式中第二项 $\frac{1}{2}m_0v^2$ 正是经典力学里熟知的物体的动能。由此可知，一个运动物体的能量 $E$ 可以分成两部分，其一是运动时才具有的动能 $T=\frac{1}{2}m_0v^2$，其二是静止时($v=0$)也具有能量 $E_0=m_0c^2$，即 $E=mc^2=m_0c^2+\frac{1}{2}m_0v^2+\cdots=E_0+T\cdots$。

物体的静能 $E_0=m_0c^2$ 是一个崭新的概念，是与物体静止质量 $m_0$ 相联系的能量，与 $m_0$ 直接成正比。这样，任何物质的质量 $m=m_0/\sqrt{1-v^2/c^2}$ 与其能量 $E$ 存在着普适关系：

$$E = mc^2 = m_0c^2 / \sqrt{1 - v^2/c^2} \qquad (3\text{-}2\text{-}7)$$

当物体的能量发生变化时，它的质量就按这一关系相应地发生变化，反之，它的质量发生了改变必定伴随其能量的变化。这一结论与目前所有的实验事实相符合。比如，人们运用质能守恒关系解释了原子核反应时质量亏损现象，发现了核内蕴藏着巨大的能量等。

### 3.2.3　物体质量的测量方法

随着科技水平的发展，质量测量技术的水平和手段在不断地提高和改进，特别是将电子技术应用到质量测定技术之中而发明的电子天平，其感量可达万分之一。目前在实验室和教学科研中常用的质量测量方法从原理上讲，主要有如下方法：

方法一：物理天平

原理：用物理天平或(各种)天平测定物体质量即天平法；

仪器：(各种)天平。

方法二：牛顿第二定律

原理：运用牛顿第二定律测质量；

仪器：验证牛顿定律的整套仪器。

方法三：利用物体的碰撞—— 动量守恒定律

原理：由动量守恒定律

$$m_1v_1 + m_2v_2 = m_1v_1' + m_2v_2'，得\frac{m_1}{m_2} = \frac{v_2' - v_2}{v_1 - v_1'}$$

仪器：气垫导轨整套仪器。

方法四：利用弹簧秤和单摆

原理：① 利用弹簧秤测物体的质量 $G$；

② 利用单摆测当地的重力加速度：$g = \frac{4\pi^2 l}{T^2}$，然后 $m = \frac{G}{g} = \frac{GT^2}{4\pi^2 l}$。

仪器：弹簧秤和单摆。

方法五：运用万有引力定律测算天体的质量

原理和仪器与方法四相同。

方法六：运用“电流天平——安培秤”测物体的质量

原理：电流天平——安培秤测量物体的质量。两个固定线圈 $C_1$ 和 $C_2$，一个活动线圈 $C_M$，通以相同的恒定电流：$mg = IlB$，如图 3-2-1 所示。则有 $m = \frac{IlB}{g}$。

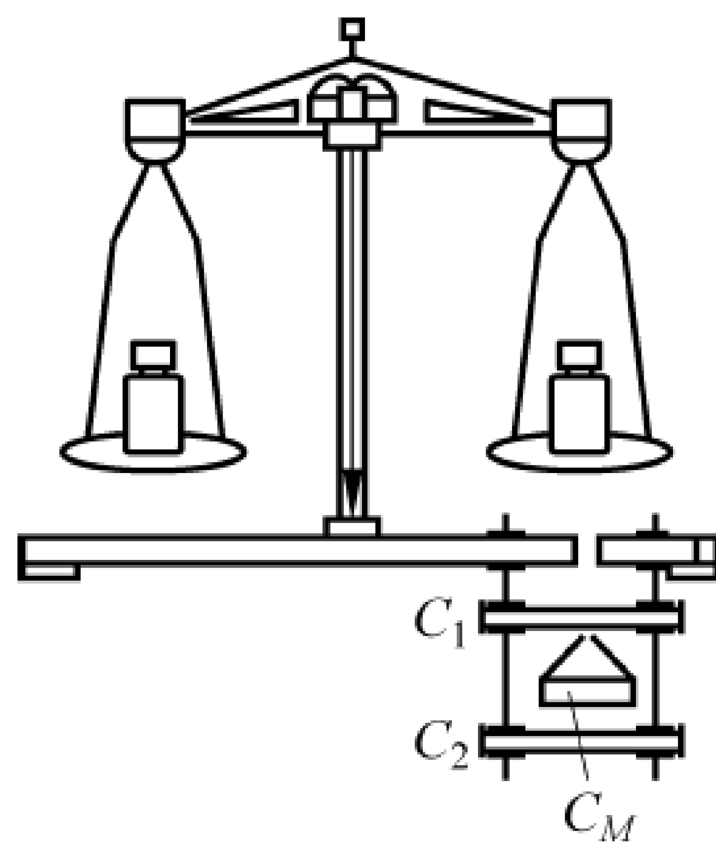


图 3-2-1　安培秤

仪器：电流天平、安培表、学生电源等。

方法七：利用匀强电场测定带电粒子的质量

原理：①带电的固体微粒所受电场力：$F = qE = q\frac{U}{d}$，得 $m = \frac{qU}{gd}$；②将带电固体微粒改为喷入的油滴即为密立根油滴实验，如图 3-2-2 所示。

仪器：密立根油滴实验仪。

方法八：利用电场和磁场(质谱仪)测定带电粒子的质量

原理：电场对带电粒子所做的功与其动能的关系为 $qU = \frac{1}{2}mv^2$，得 $v^2 = \frac{2qU}{m}$，代入

$qvB=m\dfrac{v^2}{R}$,有 $m=\dfrac{qB^2R^2}{2U}$。

仪器——密立根油滴实验仪、质谱仪,如图 3-2-3 所示。

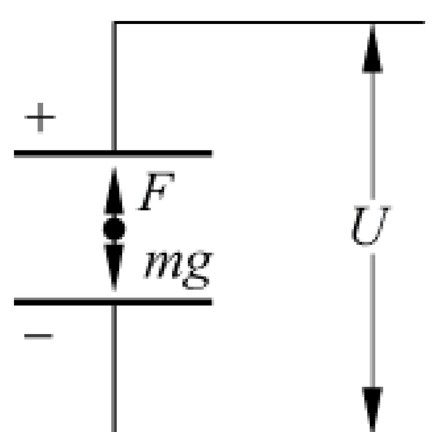



图 3-2-2 匀强电场中的带电粒子

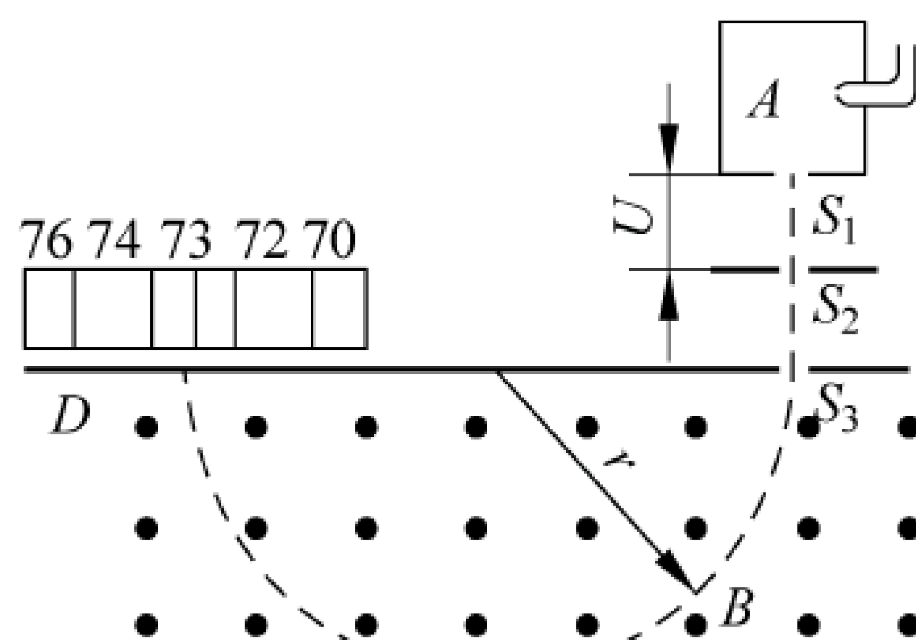



图 3-2-3 质谱仪原理

方法九:利用克拉柏龙方程测定气体的质量

原理:由克拉柏龙方程 $pV=\dfrac{m}{M}RT$,得 $m=\dfrac{pVM}{RT}$。

如果知道待测气体分子的个数 $N$,则可测气体分子的质量 $m_i=\dfrac{m}{N}=\dfrac{pVM}{NRT}$。

仪器:气体定律实验仪。

方法十:利用油膜法测定气体分子的质量

原理:用油膜法估测分子的大小。由气体的密度公式,得气体分子的质量:$m=\rho V$。

仪器:浅盘(边长 40cm 左右、深度约 3m);玻璃板(比浅盘略大);油酸酒精溶液(浓度的体积分数为 1/200 或 1/500);带针头的注射器;量杯;水彩笔;痱子粉;滴管;石膏粉等。

## 3.3 时间

现代信息社会里,人们的时间观念越来越强。

时间是一个古老、神秘而又司空见惯的概念。时间是一个普通的物理概念,但是却是一个无始无终的哲学命题。绝对时空观认为:时间是绝对、均匀流逝的自然“河流”,“逝者如斯夫”;狭义相对论认为:时间是与空间密不可分的统一体,是相对的,即与物质的运动状态有关。到目前为止,人们对时间本质的认识仍然是一知半解的。史蒂芬·霍金的《时间简史》只是人们认识和感受时间的发展史,并没有回答人们感兴趣的关于“时间本质是怎样的”问题。随着科学和技术的发展,人们在探寻时间物理和哲学本质的同时,科学测量的时间精度也越来越高。现在,以秒为时间的基本单位是国际单位制 7 个基本单位中精度最高的,三百万年误差不超过一秒。正因为它的精度高,大大促进科学技术的发展,特别是促进空间技术、计算机技术等领域的发展。

### 3.3.1 时间和时间单位的最新定义

在物理学中,物理量及其单位习惯用符号表示,故时间单位“秒”用符号 s 表示。秒的最新定义是铯 133 原子基态 2 个超精细能级之间跃迁所对应辐射的 9192631770 个周期的持续时间。

时间是物理宇宙的尺度,它给出了一个非空间的连续统一体内事物发生的次序,这表现为时间时刻;同时,时间也是由这样的次序所确定的瞬间,这表现为时刻。时刻是衡量物质运动过程事件先后顺序的物理量。如果用一根带箭头的直线表示时间轴,那么轴上的任一点就表示不同的时刻。如图 3-3-1 所示,$t_1$ 时刻早于 $t_2$ 时刻。时间间隔是两个时刻之间的

间隔长短的物理量。图 3-3-1 中的时刻 $\Delta t = t_2 - t_1$ 表示时间间隔。表 3-3-1 为已测定的时间间隔值。


图 3-3-1　时间坐标轴

时间具有连续性、单向性和序列性，而且不断向前逝去。以牛顿为代表的经典时空理论认为：时间是绝对的，与参考系无关，与空间也无关。在以爱因斯坦为代表的相对论中，时间是相对的，与参考系和空间都有密切的联系。而量子力学的建立，又对时间的连续性提出质疑，提出了最短时间间隔的观点。把时间间隔的最小单位定为时量子或“克罗农”(英文名为 chronon)。

**表 3-3-1　已测定的时间间隔**

| 名　　称 | 时间间隔/s | 名　　称 | 时间间隔/s |
|---|---|---|---|
| 宇宙年龄 | $5.0\times10^{17}$ | 中子寿命 | $9.3\times10^{2}$ |
| 地月年龄 | $1.5\times10^{17}$ | 人类正常心跳周期 | $8.0\times10^{-1}$ |
| 人类年龄 | $1\times10^{14}$ | A 调音叉周期 | $2.3\times13^{-3}$ |
| 人的寿命 | $(2\sim3)\times10^{9}$ | $\mu$ 介子半衰期 | $2.2\times10^{-6}$ |
| 免疫 B 细胞平均寿命 | $3\times10^{6}$ | 中性 $\pi$ 介子半衰期 | $0.83\times10^{-16}$ |
| 人造地球卫星周期 | $5.1\times10^{3}$ | 共振态的寿命* | $10^{-24}\sim10^{-23}$ |

* 第二和第三个基本粒子碰撞短时间结合在一起。

时间的量度一般以稳定的周期性运动为基础，以选定标准周期运动的周期的某一倍数或分数为时间单位。一般来说，时间是一种能用周期性的物理现象来观测的物理量。

由表 3-3-2 知，长期以来，地球的公转和自转被看作是最稳定的周期运动。要注意的是，时间单位虽然进行不断的定义，但这并不改变原有的意义。单位还是原来的，只是找到了一个更准确、更可靠的定义方法。更准确、更可靠的时间定义的意义在于更好地提高探索并研究这些测量仪器的稳定性。

**表 3-3-2　时间单位的演化过程**

| 名　称 | 单位制定义 | 说　　明 |
|---|---|---|
| 平太阳日 | 所谓平太阳日是假设太阳在黄道上移动的速度是均匀的，且等于视太阳运行的平均速度。这个平太阳连续二次经过某地上中天的时间间隔叫一个平太阳日。一个平太阳日的 1/86400 叫一个“平太阳秒” | 它由美国天文学家纽康在 1886 年提出。用“平太阳秒”比用“视太阳秒”精确得多，误差仅为 $10^{-7}$s，即 3 个半月差 1s。但这定义还与地球自转有关，所以带来误差在近代测量技术上还是不允许的 |
| 历法时间 | 以 1900.0 年的回归年时间长度作为制定时间单位的基础(1900.0 年是从 1900 年 1 月 0 日 12 时整开始，称为“历书”，代号为“ET”)，规定历书秒的长度为历书时的回归长度的 1/31556925.9747 | 1952 年国际天文协会第 8 届大会决议在天文历史上采用这个新时间单位。这个新定义使它等于今后起 300 年内所取的旧秒平均值，即秒的新定义没有将旧值改变，而是更加明确了。它使秒的精确度提高到 $10^{-9}$s，相当于 30 年差 1s |
| 原子钟 | 以零磁场下铯 133 原子基态的两个超精细能级($F=4\to3, mF=0\to0$)之间的跃迁所对应的辐射 9192631770 个周期的持续时间 | 这是基于原子“能级跃迁”现象以及电磁波的振荡频率有极高稳定性提出的。铯原子钟的精度由 1955 年的 $10^{-9}$s 提高到 1975 年的 $10^{-10}$s，现在已达到 $10^{-14}$s。1967 年第 13 届国际计量大会给出了这样高的时间定义 |

### 3.3.2 时间符号和时间相关量的换算关系

在米·千克·秒单位制和国际单位制中，时间的主单位都为秒。在宏观测量中，常用秒的倍数表示时间；而在微观领域中，又常用秒的分数表示时间。表 3-3-3 给出时间符号。

表 3-3-3 时间符号

| 中文符号 | 英文符号 | 国际代号 | 中文符号 | 英文符号 | 国际代号 | 中文符号 | 英文符号 | 国际代号 |
|---|---|---|---|---|---|---|---|---|
| 时间 | time | $t$ | 周期 | period | $T$ | 小时 | hour | h |
| 秒 | second | s | 分钟 | minute | min | 日 | day | d |
| 世纪 | century C | cent | 年 | year | a | 频率 | frequency | $f$ |
| 时间间隔 | time interval | $\Delta t$ | 持续时间 | duration | $t$ | 时间常数 | time constant | $\tau$ |

光年不是时间单位，而是长度单位。瞬时时间的国际代号为 $\tau$。时间间隔又常简称“时间”。秒的单位为 s，而不是 S。不同时间单位与秒的换算：

1 分(钟)＝60 秒，1 年＝365.24 日＝8766 小时＝$5.259\times10^5$ 分＝$3.156\times10^7$ 秒(对于 2 月份，平年为 28 天，闰年为 29 天)；1 世纪＝100 年＝$3.6524\times10^4$ 日＝$8.766\times10^5$ 小时＝$5.259\times10^7$ 分＝$3.156\times10^9$ 秒。在现代应用光学研究中，常用 1ms(毫秒)＝$10^{-3}$s，1μs(微秒)＝$10^{-6}$s，1ns(纳秒)＝$10^{-9}$s，1ps(皮秒)＝$10^{-12}$s，1fs(飞秒)＝$10^{-15}$s 等。

### 3.3.3 几种常见测时仪的使用方法

时间的测量仪器很多，如古代的圭表，近代的机械秒钟、石英钟，现代的原子钟等。在古代，张衡发明的水运浑仪被称为最古老的日历钟。近代发明的摆钟(伽利略在教堂发现摆的周期性。惠更斯发现了擒纵机构保持摆的摆动，在 1695 年发明第一个摆钟——自鸣钟)，大大加快了测时仪的发展。下面简单介绍几种计时装置的使用方法。

(1) 机械式秒表：机械秒表分单秒针和双秒针两种，是最常见的机械计时器。该秒表是由频率较低的机械振荡系统、锚式擒纵调速器、齿轮传动装置、有分度的表盘和指针系统组成，具有操纵秒表的启动、制动和指针回零的控制功能。它是借助摆轮游丝系统计算时间。市面上的秒表分度值有 0.1s，0.2s，0.01s，0.02s 等 4 种。要注意的是，机械式秒表的指针为跳跃式工作，所以最小刻度以下的估计值无意义。

(2) 数字毫秒计：它是用数字显示的电子计时仪器。它用石英晶体振荡分频作为时基脉冲，能产生稳定在 10kHz 的电脉冲，即每秒钟内准确产生 10000 个脉冲，然后通过这些脉冲在开始计数和停止计数的时间间隔去推动计数器计数。一个脉冲计一个数，任何两个相邻脉冲时间间隔为 $10^{-4}$s，即 0.1ms。所以通过计数器所计的数，就很容易测出从“计”到“停”的这段时间长短，并在数码显示器装置上直接显示出来。

(3) 原子钟：原子钟利用了微观粒子超精结构能级之间的跃迁，其跃迁产生高准确度和高稳定度的周期振荡现象。由石英振荡器输出一定的频率，经过受控振荡器得到稳定性高的信号，再经放大、分频、门控电路和液晶显示器，显示所测时间。原子钟按其工作原理方式可分为自激式和非自激式两种。

此外，常用的计时方法还有示波器、高压脉冲发生器、频闪装置等。这可根据测量的精确度选用合适的计时仪。有兴趣的读者可以参考专门的文献。

### 3.3.4 计量的新应用

任何计时装置都由能产生稳定的周期运动系统和维持这一运动的能源及记录时间机构

组成。关系式 $T=1/f$($T$ 为周期,$f$ 为频率)表明一个稳定的计时仪也是一个稳定的频率计。同时,时间的精度也表示频率的精度。最近,美国公司研制的一台新 GPS 校频接收机,通过给出稳定和准确的 10MHz 频率来定准一周时间,可精确到 $5\times10^{-12}$s。

获 1999 年诺贝尔化学奖的科学家泽维尔(Ahmed H. Zewail,1946—　)就是研究毫微微秒光谱而获奖的。他设计的摄影装置可以将一毫微微秒的活动定格。现在,全世界的科学家都在运用毫微微秒光谱技术来研究气态、固态的表面或内部的化学变化过程。

总之,1967 年基于铯原子跃迁而产生新的秒定义,完全改变了时间与频率计量的概念特征。由于在时间基准、时间同步及新的频率标准方面的研究成果,法国计量局时频实验室已成为主导实验室之一。

### 3.3.5　时间的发展方向

随着科学研究的不断深入和测量技术的迅速发展,时间的计时装置也正朝高稳定和高精度的方向发展。时间单位的定义不仅关系到计时仪的改进,而且也影响到其他 6 个国际单位的基准定义。总的说来,从时间单位的发展过程和应用情况来看,它的方向可能朝以下几方面发展:

(1) 人们对时间的定标,由自然现象转向物质结构(铯原子),由现实性转向理想化假设,由误差大向精度高转化。

(2) 时间的周期测量,不仅与频率相关,还和热平衡、温度系统等有关。

(3) 时间测量将考虑太阳系参与银河系的运动,时钟受到银河系的引力影响,人们企图寻找新的坐标参考系来定义时间,使它与相对论的时空概念不矛盾。

(4) 将出现有关未来工程的相对论。时针技术的改进有可能得到更准确的定时、测距、导航和通信系统。为了使现代化的系统能正确工作,有 3 种相对效应必须考虑,它们是:时间膨胀、SAGNAC 效应和引力频移。

(5) 人们对反馈时间更感兴趣,不仅致力于研究时间的"倒计时",更注重研究时间的反馈信息。

**附录**:关于史蒂芬·霍金和他的名著《时间简史》

史蒂芬·霍金(Stephen William Hawking,1942—2018)是英国著名物理学家,被誉为是继爱因斯坦之后最杰出的理论物理学家之一。他提出宇宙大爆炸自奇点开始,时间由此刻开始和黑洞最终会蒸发等观点。他先后毕业于牛津大学和剑桥大学,并获剑桥大学哲学博士学位。他因患"渐冻症"(肌萎缩性侧索硬化症),被禁锢在一把轮椅上达 40 年之久却身残志不残,克服了残废之患而成为国际物理界的超新星。他不能书写,甚至口齿不清,但他的研究超越了相对论、量子力学、大爆炸等理论。他编著的《时间简史》是一本人人都能读懂的科学经典名著,从 1988 年首版发行至今,已经发行了 40 多种语言,销量超过 2500 万册。

《时间简史》目录:

第 1 章　我们的宇宙图像

第 2 章　空间和时间

第 3 章　膨胀的宇宙

第 4 章　不确定性原理

第 5 章　基本粒子和自然的力

第 6 章　黑洞

第 7 章　黑洞不是这么黑的

第 8 章　宇宙的起源和命运
第 9 章　时间箭头
第 10 章　虫洞和时间旅行
第 11 章　物理学的统一
第 12 章　结论

## 3.4　温度

温度是热学中最重要的基本概念之一，也是国际单位制 7 个基本量中在热学领域内特有的物理量。温度概念的建立和测量是建立在热平衡基础之上的。伴随着人们对经典物理现象及概念的逐步深入认识和研究，对温度的理解、界定和测量方法经历了一个漫长的发展变化过程。直到 1960 年第十一届国际计量大会通过的国际单位制(SI)，才将“热力学温度”(thermodynamic temperature)确定为国际单位制的 7 个基本量之一。规定热力学温度(简写为 $\Theta$，但是习惯上仍然用大写英文字母 $T$ 表示)的单位名称为“开尔文”，简称“开”，其国际单位代号为“K”，其文字定义为：“1 开等于水的三相点的热力学温度的 1/273.16。”热力学温度的实现——热力学温标，其零点，即绝对零度，记为“0K”。由于水的三相点在摄氏温标上为 0.01℃，所以 0℃＝273.15K。

温标概念在实际测温中具有重要意义，本节列举了不同类别的温标，较详尽地阐述目前国际上用得较多的温标，即华氏温标、摄氏温标、热力学温标等，并说明温标制定是通过一系列理论的推证和实际的测量的复杂过程而得出的。最后从回顾温度测量历史发展的角度，简要地总结介绍几种常用的测温方法。

### 3.4.1　温度的概念

物体的“冷热程度”最初是来源于人们的主观感觉，热的物体温度高，冷的物体温度低。习惯用冷、温、热、烫等文字定性地来描述，但是这种表示方法模糊、粗糙、不准确。在生活、生产和技术中，更需要对物体冷热程度的精确描述——定量化描述。两个物体发生热接触后，经过足够长的时间后，它们的状态不再发生变化，而达到一共同的平衡态。这时两个物体达到了热平衡，如果两个物体各自与第三个物体达到热平衡，它们彼此必处在热平衡。这个事实称为热平衡定律。互为热平衡的物体必有一个共同的物理性质，这个性质保证它们在进行热接触时达到热平衡。经验告诉人们，这个性质就是它们的冷热程度——温度。热平衡定律不仅给出了温度的概念，而且指明了比较温度的方法。由于互为热平衡的物体有共同的温度，我们比较两个物体的温度时，不需要令两个物体直接接触，只需取一个标准的物体分别与这两个物体进行热接触就行了。这个作为标准的物体就是温度计。当然，要定量地确定温度的数值，还必须对不同的冷热程度给以数值的表示，即确定温标。

经典物理学建立温度概念的理论基础是热力学第零定律：在不受外界影响的情况下，如果系统 A、B 分别同时与系统 C 处于热平衡，则 A、B 必然处于热平衡。因此，温度概念的建立和定量测量都是以热平衡现象为基础的。两个物体发生热接触(不一定是直接接触)后，经过足够长的时间后，它们的状态不再发生变化，而达到一共同的平衡态。这时两个物体达到了热平衡。如果两个物体各自与第三个物体达到热平衡，它们彼此必处于热平衡。因此，互为热平衡的物体必有一个共同的物理性质，这个性质保证它们在进行热接触时达到热平衡。由热力学第零定律可知，一切互为热平衡的物体都具有某一相同的宏观物理性质，

表征这一宏观物理性质的量就称为温度。

一般情况下，定量描述一个热力学系统与其他热力学系统处于热平衡状态，即处于共同的物理性质，用热平衡的状态函数表示：$T=f(p,V)$——绝对温度的宏观定义式，其中 $T$ 即是系统的绝对温度，$p$ 是系统的压强，$V$ 是系统的体积。

热力学第零定律不仅给出了温度的概念，而且指明了比较和测量温度的方法。由于互为热平衡的物体具有共同的态函数——温度，我们比较两个物体的温度时，不需要令两个物体直接接触，只需取一个标准的物体分别与这两个物体进行热接触就行了。这个作为标准的物体通常用温度计来实现。能够作为温度计的物质，必须具有三个要素：测温物质的温度属性、选定零点和合理分度。要定量地确定温度的数值，还必须对不同的冷热程度给以数值的表示，即确定温标。温标是温度数值表示方法——依赖于测温物质三个要素。

随着经典物理学的发展，关于温标的实现，经历了四个过程：即经验温标、理想气体温标、热力学温标(绝对温标)和国际实用温标。

### 3.4.2　温标及建立温标的几个要素

#### 1. 三相点

三相点亦称“三态点”。各种稳定的纯物质处于固态、液态、气态三个相(态)平衡共存时的状态，叫作该物质的“三相点”。该点具有确定的温度和压强。

物态叫作“相”，通常物质是以三种形态存在，即固态、液态、气态，也可称为固相、液相、气相。物态的变化常叫作相变。或者说，在某一系统中，物理性质均匀的部分亦称为相。相与相间必有明显可分的界面。例如，食盐的水溶液是一相，若食盐水浓度大，有食盐晶体，即成为两相。水和食油混合，是两个液相并存，而不能成为一个相。又如水、冰和汽三相共存时，其温度为 273.16K(0.01℃)，压强为 $6.106\times10^2$Pa。由于在三相点物质具有确定的温度，因此用它来作为确定温标的固定点比选汽点和冰点具有优越性，所以三相点这个固定温度适于作为温标的基点，现在都以水的三相点的温度作为确定温标的固定点。表 3-4-1 列出几种物质三相点的数据。

**表 3-4-1　几种物质三相点的数据**

| 物　　质 | 温度/K | 压强/Pa |
|---|---|---|
| 氢 | 13.84 | 7038.2 |
| 氘 | 18.86 | 17062.4 |
| 氖 | 24.57 | 43189.2 |
| 氮 | 63.18 | 125330.2 |
| 二氧化碳 | 216.55 | 517204.0 |
| 水 | 273.16 | 610.5 |

#### 2. 几种温标简介

温标是温度的标尺，在历史上，人们曾经用过许多温度单位，温标热学发展史上曾出现过华氏温标、列氏温标、兰氏温标、摄氏温标、气体温标、热力学温标和国际温标等。

摄氏温标是由瑞典天文学家安德斯·摄耳修斯(A. Celsius，1701—1744)在 1742 年提出的。规定在标准大气压下，冰的熔点为 0 摄氏度，水的沸点为 100 摄氏度，中间划分 100 等份，每等份为 1 摄氏度(℃)。

华氏温标是由德国物理学家华伦海特(Daniel Gabriel Fahrenheit，1686—1736)大约在

1710 年提出的，规定在标准大气压下，冰的熔点为 32 度，水的沸点为 212 度，中间划分 180 等份，每等份为 1 华氏度。以℉表示，与摄氏温标℃的换算关系为：℉=9/5℃+32。华氏温标至今还在英、美等国的民间流行。

列氏温标是由列奥缪尔(R. A. F. Reaumur)于 1730 年提出的，规定水的冰点为零度，水的沸点为 80°，以°R 表示，与摄氏温标℃的换算关系为：℃=1.25×°R+273.15。列氏温标在德国曾一度流行。

兰氏温标：该温标是美国工程界使用的一种温标。开氏温标以水的三相点为 273.16K，兰氏温标以 273.16K 作为 491.688°R。它们都是从绝对零度起算，所以也叫绝对温标。华氏温度 $t_F$ 与兰氏温度 $T_R$ 的关系是：$t_F=T_R-459.67$。

$$兰氏温度=(摄氏温度+273.15)\times 9/5$$

气体温标是以气体温度计标定温度所构成的温标。由于气体温度计的复现性较差，国际间又协议定出国际实用温标，以统一国际间的温度值，为了定出的温度尽可能接近热力学温度，国际实用温标几经变革。以上几种温标都是依据某一物质的测温特性而确定的温标称为经验温标。而把不依赖于任何具体物质的测温特性的温标称为绝对温标。

绝对温标，也称为热力学温标，是由汤姆孙(William. Thomson，即开尔文，Lord. Kelvin，1824—1907)提出的。汤姆孙在法国著名的实验物理学家勒尼奥(H. V. Regnault，1810—1878)的实验室里工作时，第一次读到了克拉珀龙(Clapeyron，1799—1864)阐述卡诺(Sadi Carnot，1796—1832)热动力理论的文章，对卡诺理论的威力留有深刻的印象。首先引起汤姆孙注意的，是可以通过卡诺的热机确定温度，因为卡诺热机与工作物质无关，这样定出的温标比根据气体定律建立的温标有许多优越的地方。1848 年，汤姆孙在题为《基于卡诺的热动力理论和由勒尼奥观测结果计算所得的一种温标》的论文中写道："按照卡诺所建立的热和动力之间的关系，热量和温度间隔是计算从热获得机械效果的表达中唯一需要的要素，既然我们已经有了独立测量热量的一个确定体系，我们就能够测量温度间隔，据此对绝对温度差做出估计。"汤姆孙还对这个温标作了如下的说明："所有度数都有相同的值，即物体 A 在温度 $T$，有一单位热量由物体 A 传到温度为 $T_1$ 的物体 B，不论 $T$ 值多大，都会给出同样大小的机械效果。这个温标应正确地称为绝对温标，因为它的特性与任何特殊物质的物理性质是完全无关的。"1854 年，汤姆孙和焦耳(J. P. Joule，1818—1889)联合发表了《运动中流体的热效应》一文，其中专门有一节题为"根据热的机械作用建立的绝对温标"，他们定义绝对温标为：$T\equiv J/\mu$(其中 $J$ 为热功当量值，$\mu$ 为卡诺系数)，可得：$T=t+1/E$。如果取 $E=0.003665$，则 $T=272.85+t$。考虑到其他因素，他们得到的修正结果为：$T=273.3+t$，这就是绝对温标和摄氏温标的关系。绝对温标的建立对热力学的发展有重要意义。汤姆孙的建议很快就被人们接受，1887 年，绝对温标得到了国际公认。早在 1848 年，汤姆孙确定绝对温标时，对绝对零度作了如下说明："当我们仔细考虑无限冷相当于空气温度零度以下的某一确定的温度时，如果把分度的严格原理推延足够远，我们就可以达到这样一个点，在这个点上空气的体积将缩减到无，在刻度上可以标－273℃，所以空气温度计的－273℃是这样一个点，不管温度降到多低都无法达到这点。"绝对零度不可能达到，在物理学家的观念中似乎早已隐约预见到了。但是这样一条物理学的基本原理，却是又过了半个多世纪，到 1912 年才正式提出来。1912 年，德国物理化学家能斯特(Walther Hermann Nernst，1864—1941)在他的著作《热力学与比热》中，将"热学新理论"表述成："不可能通过有限的循环过程，使物体冷到绝对零度。"这就是绝对零度不可能达到定律，也是热力学第三

定律通常采用的表述方法。

国际实用温标是一个国际协议性温标，它与绝对温标相接近，而且复现精度高，使用方便。国际计量委员会在十八届国际计量大会第七号决议授权予1989年会议通过了1990年国际温标ITS-90。国际温标ITS-90同时定义国际开尔文温度（符号为T90）和国际摄氏温度（符号为t90）。国际温标ITS-90的通则ITS-90由0.65K向上到普朗克辐射定律使用单色辐射实际可测量的最高温度。ITS-90是这样制订的，即在全量程中，任何温度的T90值非常接近于温标采纳时$T$的最佳估计值，与直接测量热力学温度相比，T90的测量要方便得多，而且更为精密，并具有很高的复现性。ITS-90的定义，第一温区为0.6～5K之间，T90由$^3$He和$^4$He的蒸气压与温度的关系式来定义。第二温区为3.0K到氖三相点（24.661K）之间，T90是用氦气体温度计来定义。第二温区为平衡氢三相点（13.033K）到银的凝固点（961.8℃）之间，T90是由铂电阻温度计来定义。它使用一组规定的定义固定点及利用规定的内插法来分度。银凝固点（961.78℃）以上的温区，T90是按普朗克辐射定律来定义的，复现仪器为光学温度计。

理想气体温标是以气体作为测温物质，利用理想气体物态方程中体积（或压强）不变时压强（或体积）与温度呈正比例关系所确定的的温标：$T(p)=\alpha p$，或$T(V)=\beta V$。选水的三相点为固定点，规定水的三相点温度为273.16K，水的饱和蒸气压为610.5Pa。因此可以得到理想气体温度计测温泡中气柱在水的三相点时的压强$p_{\mathrm{tr}}$和体积$V_0$。理想气体在体积$V_0$一定的情况下，其物态方程为$pV_0=\nu RT(p)$，则理想气体柱在水的三相点时的压强$p_{\mathrm{tr}}=\dfrac{\nu R}{V_0}\times 273.16\mathrm{K}$，则理想气体温度计压强为$p$时所测出的温度为

$$T(p)=273.16\mathrm{K}\cdot\lim_{p_{\mathrm{tr}}\to 0}\frac{p}{p_{\mathrm{tr}}}\quad（体积不变）\tag{3-4-1}$$

理想气体温标用气体温度计来实现，但读数与气体的个性无关。受气体共性限制，在气体液化点以下及高温下，理想气体温标不适用。由于氦的液化温度最低，且不易在金属（铂）中扩散，所以氦温度计具有一定的优越条件。

国际实用温标，从准确与实用出发，在1927年第七届国际计量大会上决定采用国际温标。由于科学技术不断地发展，工业生产上的需要，国际温标不断修改，目前所采用的国际实用温标，是1968年国际计量委员会对1948年国际实用温标（1960年修正版）作了重要修改而建立的。1968年国际实用温标选取的方法，是根据它所测定的温度可紧密接近热力学温度，而其差值应在目前测定准确度的极限之内。1968年国际实用温标在国际实用开尔文温度和国际实用摄氏温度之间是用符号$T_{68}$和$t_{68}$来加以区分的，$T_{68}$和$t_{68}$之间的关系是：$t_{68}=T_{68}-273.15$。$T_{68}$和$t_{68}$的单位如在热力学温度$T$和摄氏温度$t$中一样仍为开尔文（符号K）和摄氏度（符号℃）。常用的换算公式是$T=t+273.15$。

热力学温标，亦称“开尔文温标”“绝对温标”。它是建立在热力学第二定律基础上的一种和测温物质无关的理想温标，它具有完全不依赖测温物质的性质。1927年第七届国际计量大会曾采用其作为基本的温标。1960年第十一届国际计量大会规定热力学温度以开尔文为单位，简称“开”，用K表示。根据定义，1开等于水的三相点的热力学温度的1/273.16。由于水的三相点在摄氏温标上为0.01℃，所以0℃＝273.15K。热力学温标的零点，即绝对零度，记为“0K”。热力学温标，按照国际规定是最基本的温标，它只是一种理想温标。理想

气体温标由于在它所能确定的温度范围内等于热力学温标，所以往往用同一符号 $T$ 代表这两种温标的温度。在理想气体温标可以实现的范围内，热力学温标可通过理想气体温标来实现。

随着国际上科学技术的交流，现在用得最多的是瑞典摄尔修斯提出的摄氏温度。我们日常生活中说的某某物体的温度或气温多少度都是指的多少摄氏度。

## 3.5 电流强度

电流强度又称电流，是一个重要的基本物理量，在国际单位制中，电流的单位安培(A)是 7 个基本单位之一。选用电流强度作为基本物理量是有原因的，电流遍及整个庞大的电磁学体系，可以这样说，没有了电流，就等于没有了整个电磁学大厦的存在。而且，电流的测量技术是电磁测量技术的基础，电磁学中包含电流的定义公式较多，导出其他电学量和磁学量的单位比较方便。

1948 年，第九届国际计量大会责成国际计量委员会，正式征询所有国家科学技术界与教育界的意见，以制定一种能为所有米制公约签字国都乐意采用的实用单位制。大会推荐 MKS(米千克秒)制和绝对实用制的一个电学单位。

**1. 电流的形成**

电荷流动形成电流。在宏观范围内，电流就是大量电荷的定向运动。要产生电流，一方面必须存在可以自由运动的电荷，即载流子，另一方面必须有迫使电荷作定向运动的某种作用。由于导体对载流子的定向运动有阻力(这种阻力分别来自载流子与晶格或其他中性分子的碰撞)，要维持电荷的定向运动，这种作用是必不可少的。

在多数的情况中，载流子是电子或某种带电微粒，如正、负离子。迫使电荷作定向运动的作用则是多种多样的，有机械作用、化学作用、电作用等。金属导体中电流的载流子是自由电子。金属中存在大量自由电子，当金属处在电场中时，自由电子因电场力而作定向运动，从而形成金属中的电流，如图 3-5-1 所示。由于电子的质量很小，金属中电流不会引起宏观上可观察到的质量迁移。

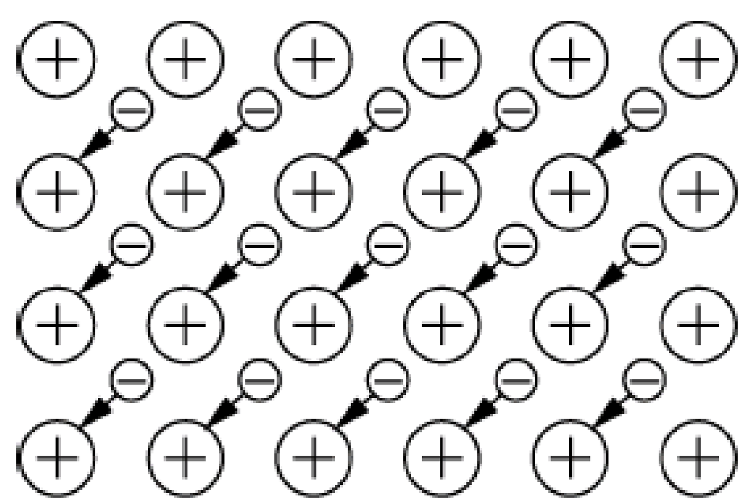


图 3-5-1 金属中自由电子的定向运动形成电流

酸、碱、盐等电解质溶液中电流的载流子是正离子和负离子。当电解质溶液处在电场中时，正、负离子因受电场作用而分别向相反的方向作定向运动，从而形成电解质溶液中的电流。从电量迁移的角度看，正电荷向某一方向运动与负电荷向相反的方向的运动所产生的效果是相同的。电解质溶液中的电流会引起质量的迁移，一般还伴随化学反应。

半导体材料中的载流子是电子(导带中)和带正电的空穴(满带中)，电子或空穴在电场作用下形成半导体中的电流。半导体中载流子的密度和定向运动的速度与温度、光度等因素密切相关。

通常，气体中没有可以自由移动的电荷，故气体没有导电性，是良好的绝缘体。但是，紫外线、X 射线、宇宙线以及火焰等所谓电离剂会使气体分子电离，产生电子和正、负离子，从而使气体具有导电性。在电场作用下，电子和离子的定向运动就形成气体中的电流。电子在电场的作用下可以获得很大的动能，当它们与中性分子碰撞时有可能使中性分子电离而产生新的载流子。

真空中没有自由电荷,故在一般情况下真空中不会有电流。

金属内部的自由电子可以在金属内部自由运动,但是它们很难进入真空中形成电流。不过随着金属温度的升高,动能大的电子增多,当金属达到灼热时,动能大的电子会很多,从而有大量电子从金属中溢出,这就是热电子发射。热电子发射使真空中出现大量的载流子,在外电场的作用下形成真空中的电流。微观粒子具有贯穿势垒的隧道效应,即使金属的温度不高,电子仍有一定的概率贯穿势垒进入真空,从而可在特定的条件下使真空中形成微弱的隧道电流。

**2. 电流强度与电流密度**

根据《国际通用计量基本名词》,"量"定义为:现象、物体或物质的可以定性区别和定量确定的一种属性。由此定义可知,被研究的对象可以是自然现象,也可以是物体本身。而电流强度作为国际单位制(SI)基本物理量,研究的就是大量电荷的定向运动的现象。

由于电子服从量子力学的规律,即使处于绝对零度附近,金属中的自由电子仍必须分布在一系列能量不同的状态上,因而电子不规则运动的平均速率仍非常大,其数值约为 $10^6$ m/s。但是,电子的平均速度为零,所以,电子的不规则运动并不引起宏观上的电流。

由于负电荷向某一方向的定向运动所引起的电量迁移与等量的正电荷向相反的方向作定向运动所引起的电量迁移等效,加之传统习惯,即使在很多情况下实际的载流子是带负电的电子,但在研究电流时都规定带正电的载流子的定向运动方向作为电流的方向。

考察某种导体材料,其载流子密度(即单位体积内载流子数)为 $N$,载流子所带电量为 $q$,其定向运动的速度为 $\boldsymbol{u}$,则在 $\Delta t$ 时间内,通过任一面元 $\Delta s$ 迁移的电量为

$$\Delta Q=(u\Delta t\Delta s\cos\theta)Nq$$

其中 $\theta$ 为 $\Delta s$ 的法线方向与 $u$ 的夹角。单位时间内通过任一面积迁移的电量表示为通过该面积的电流的大小,因此通过 $\mathrm{d}s$ 面的电流为

$$\mathrm{d}I=\frac{\mathrm{d}Q}{\mathrm{d}t}=\frac{(u\,\mathrm{d}s\,\mathrm{d}t\cos\theta)Nq}{\mathrm{d}t}=Nqu\,\mathrm{d}s\cos\theta=Nq\boldsymbol{u}\cdot\mathrm{d}\boldsymbol{s}\tag{3-5-1}$$

令 $\boldsymbol{j}=Nq\boldsymbol{u}$,则

$$\mathrm{d}I=\boldsymbol{j}\cdot\mathrm{d}\boldsymbol{s}\tag{3-5-2}$$

其中 $\boldsymbol{j}$ 称为电流密度,它是矢量,方向与载流子定向运动速度的方向相同,大小等于单位时间内通过垂直于载流子定向速度方向的单位面积的电量。

如果电流是由几种载流子定向运动形成的,则每一种载流子的定向运动对电流的贡献都由式(3-5-1)给出,故有 $\mathrm{d}I=\left(\sum_i N_i q_i \boldsymbol{u}_i\right)\cdot\mathrm{d}\boldsymbol{s}$;而电流密度 $\boldsymbol{j}=\sum_i N_i q_i \boldsymbol{u}_i$ 通过任意曲面的电流:

$$I=\int_s \boldsymbol{j}\cdot\mathrm{d}\boldsymbol{s}=\int_s j\cos\theta\,\mathrm{d}s\tag{3-5-3}$$

电流是单位时间内通过某一曲面的总电量,而电流密度则反映了空间各点电流的分布情况。例如一般地,通过各截面上的电流相等,但是各个截面的电流密度却可能不同,同一截面上不同处的电流密度也不相同,如图 3-5-2 所示。

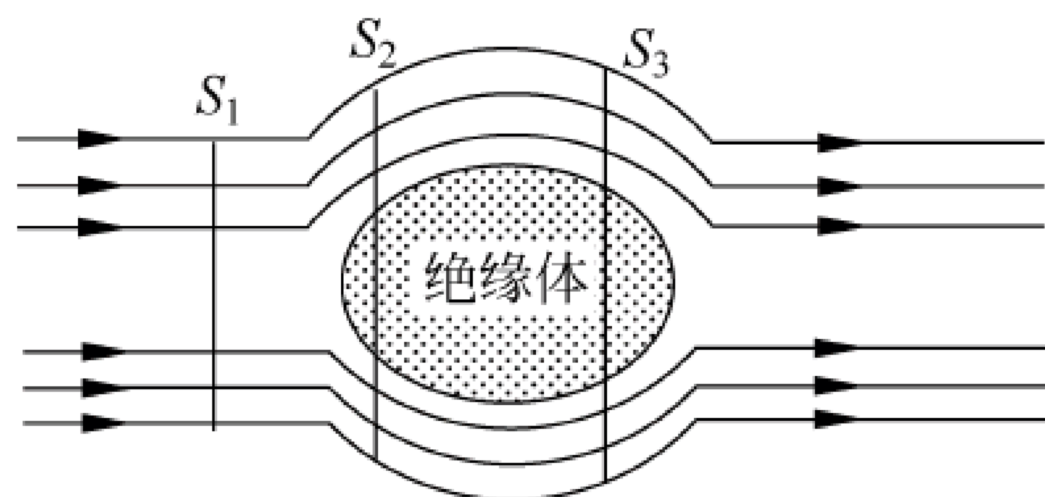



图 3-5-2　电流通过三个不同截面的电流密度

**3. 安培的定义及其演变**

电流的物理意义是单位时间内通过导体横截面

的电量,即 $I=Q/t$。如果在 1 秒(s)内通过导体横截面的电量为 1 库仑(C),导体的电流就是 1 安培(A)。我们知道一个电子的电荷约等于 $1.6\times10^{-19}$C,如果能测出 1s 内通过导体某一横截面的电子数,就可以复现安培单位。但是实际上,利用目前的技术很难实现。因此 100 多年来各国的科学家均在其当时所能实现的技术条件下,利用其他各种不同的方法来复现安培。

(1) 1893 年以前安培的复现方法

19 世纪中叶,随着工业化的发展,电磁学的研究和应用有了明显的进展。人们开始对力学、电学和磁学的各物理量单位的选择进行了大量的研究讨论。在那段时间里,虽然已经出现了通过测定在单位时间内电解银的数量来测量电流的银电解电量计,但对安培还没有专门的定义。实际上安培是通过欧姆定律由电压和电阻标准来复现和保存的,复现单位量值的准确度只有 $10^{-4}$ 量级。

(2) 1908 年对安培的定义

1893 年在芝加哥召开的第四次国际电学会议上确定采用以法拉第常数为基础的银电解电量计为国际安培基准器。考虑到用实物基准复现的单位量值与它们对应的绝对单位会有较大差别,于是在 1903 年的一次国际会议上,提出了建立基于实物基准的国际实用单位的建议。

1908 年,在伦敦举行了有 24 个国家官方代表参加的电学单位和国际标准国际会议。会上决定使用国际实用单位,以电解银的电流来定义"国际安培"。

1 国际安培为通过在按规定配制的硝酸银水溶液中以每秒 0.00111800 克的速率电解银时的恒定电流。

这段时间里,各国科学家还开展了电单位的绝对测量,即以长度、时间和质量等基本单位来复现电单位。随着绝对测量正确度的提高,发现当时使用的国际实用单位与绝对实用单位之间存在万分之几的差别。于是,在电磁学领域内一场较大的变革又开始酝酿了。

(3) 1948 年对安培的定义——SI 安培的定义

1935 年 CIPM 的电学咨询委员会(CCE)根据 1933 年第八届 CGPM 的决定,准备从 1940 年 1 月 1 日起重新改用绝对实用单位,并公布了一个临时换算关系。但是由于第二次世界大战的爆发,这个决议没有实现。战后,于 1946 年召开的 CIPM 会议上,根据当时绝对测量的结果,重新确定了国际实用单位与绝对实用单位之间的换算关系,并决定自 1948 年 1 月 1 日起开始实施。这是现代计量史上电单位的第一次大的改值。新的换算关系为

1 绝对安培=1.00015 国际安培

1 绝对欧姆=0.99951 国际欧姆

1 绝对伏特=0.99966 国际伏特

1948 年第九届国际计量大会正式批准通过的 1946 年国际计量委员会的决议,正式采用 *MKSA* 制,即在以米千克秒制代替厘米克秒制的同时,增加第四个基本单位安培。大会在重新定义安培时,为了使以新定义的安培为基础的电磁力单位和以米·千克·秒制为基础的机械力单位相等,引入了一个系数 $\mu_0$,$\mu_0=4\pi\times10^{-7}\mathrm{N/A^2}$。安培的新定义是:安培是电流的单位。在真空中,截面积可忽略的两根相距 1 米的无限长平行圆直导线内通以等量恒定电流时,若导线间相互作用力在每米长度上为 $2\times10^{-7}$ 牛顿,则每根导线中的电流为 1A。上述定义的数学式为

$$1\mathrm{A}=\sqrt{\frac{2\pi\times2\times10^{-7}\mathrm{N}}{\mu_0}}\quad(\text{式中 }\mu_0\text{ 为真空磁导率})$$

图 3-5-3 为该定义的示意图。

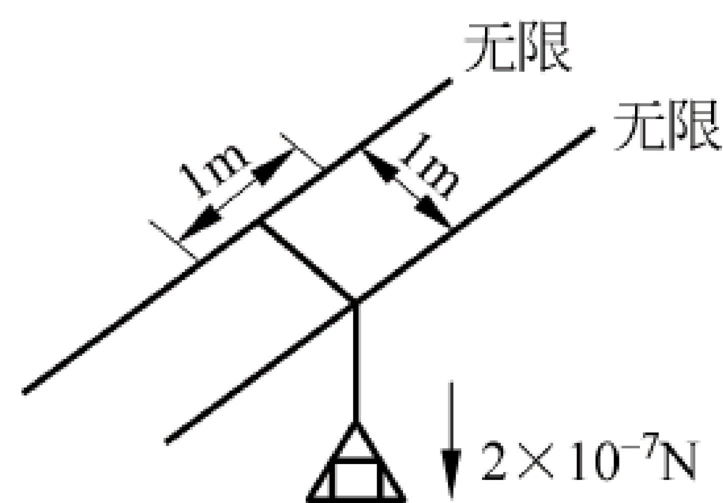


图 3-5-3　安培定义的示意图

由物理实验可知，两条互相平行的直导线上，如果有相同方向的电流通过，那么这两条互相平行的直导线之间就会产生吸引的力；当这两条互相平行的直导线有相反的电流通过时，就会产生排斥力。当导线间的距离等条件一定时，两导线间的排斥力（或吸引力）是与电流的大小成正比的。因此，可以利用这种力的大小测定电流的大小。国际单位制就是根据这个道理来确定安培为电流单位的。

**4. 安培的复现——电流的绝对测量**

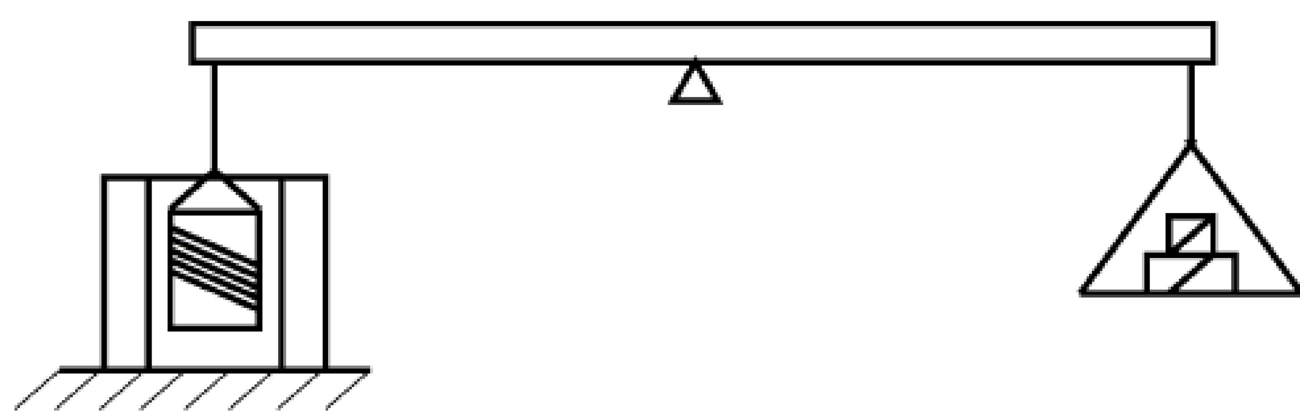
基本单位安培定义了以后，各个国家用实物基准所保存的安培量值是否与定义值（亦称为 SI 值）相等，必须通过实验来测定，也就是说按照定义来复现安培并与实物基准保存的安培量值相比较。在 $\mu_0$ 被作为精确常数规定之后，安培按其定义通过力学量的基本单位（m、kg、s）来复现，称之为电流的绝对测量。

自 20 世纪 20 年代至今，各国科学家进行了各种电流的绝对测量方法的研究，有的因其准确度低而被逐渐淘汰，有的至今仍在采用。

下面对其中的两种电流的绝对测量方法予以简单的介绍。

（1）电流天平法

20 世纪 40—60 年代，一些主要的工业国家如美、英、苏联等以及 20 世纪 60—70 年代的东欧一些国家都采用电流天平法对电流进行了绝对测量。图 3-5-4 给出了电流天平原理图，此方法巧妙地用了一个可移动螺线管线圈置于一个固定螺线管线圈中间，近似地实现了安培定义中两根无限长且相互平行的直导线的条件，同时，用天平称出两线圈通电时的相互作用。实际上这个装置与安培定义中所要求的条件还不完全相符，因此，选取麦克斯韦电磁理论中的两通电回路相互作用的能量公式 $W=I_1I_2M$，式中 $I_1$、$I_2$ 分别为两通电回路中的电流；$M$ 为两通电回路之间的互感。对此式经过处理可得到：$\Delta F=mg=\mathrm{d}W/\mathrm{d}x=I_1I_2\mathrm{d}M/\mathrm{d}x$。


图 3-5-4　电流天平原理图

一般都使可动线圈与固定线圈串联，即 $I_1=I_2=I$，$\mathrm{d}M/\mathrm{d}x$ 可根据线圈的几何尺寸求出。在测出作用力 $\Delta F$ 后，即可计算出电流的大小

$$I=\left(\frac{\Delta F}{\mathrm{d}M/\mathrm{d}x}\right)^{\frac{1}{2}} \tag{3-5-4}$$

线圈中通过的电流一般为 1 安培，两线圈的相互作用力很小，单是测力这一项就使电流天平的测量不确定度限制在 $10^{-6}$ 量级，加上线圈几何尺寸及它们之间的相对位置的测量都不可能达到很高的准确度，因此，几十年来，电流天平法复现安培的不确定度一直徘徊在百万分之一范围内。目前此方法已被逐渐淘汰。但应该肯定，在电学计量史上电流天平对提高电单位复现准确度曾起过重要作用。

（2）核磁共振法

20 世纪 60 年代以来，一种复现安培的新方法——核磁共振法在一些国家中开始研究和应用。它是通过在弱磁场和强磁场中分别测定基本物理常数——质子回旋比 $\gamma_p$，利用

$\gamma_p$(强)与电流成正比、$\gamma_p$(弱)与电流成反比的特点,将两者相结合,计算得到保存安培与绝对安培之比 $K_A$(=$A_{LAB}$/A)(由于篇幅所限,这里不具体介绍弱磁场核磁共振和强磁场核磁共振测定质子回旋比 $\gamma_p$ 的方法)。

我国,还有美、英、日等国进行过 $\gamma_p$(弱)测量,测量准确度最高的为美国国家标准技术研究所,1988 年其测量不确定度为 $0.11\times10^{-6}$。英、苏联和德国还进行过 $\gamma_p$(强)的测量,英国国家物理研究所的测量准确度最高,1974 年所发表数据的不确定度为 $1.0\times10^{-6}$,但他们和苏联后来都中止了这一工作。

**5. 安培的保存**

由于电功率天平法(篇幅所限,这里不作介绍)可能达到 $10^{-8}$ 甚至 $10^{-9}$ 量级的不确定度,而强场 $\gamma_p$ 测量的不确定度很难显著减少,因此,核磁共振法在绝对测量电流方面的作用将会下降。但由于 $\gamma_p$ 是一个重要的基本物理常数,继续提高 $\gamma_p$,特别是 $\gamma_p$(弱)的测量准确度对精确确定某些基本物理常数如精细结构常数 $\gamma_p$ 等仍是相当重要的。

(1) 1990 年以前安培的保存——电压和电阻的实物基准

在 1990 年国际上正式启用自然基准之前,电学单位是通过实物基准来复现和保存的。在电学计量领域里最重要的实物基准是电阻基准和电动势基准。电流和电功率没有实物基准,它们是通过欧姆定律和电功率公式由电阻和电动势基准一起来复现和保存,目前各国大多数仍采用锰铜合金电阻和惠斯通饱和标准电池分别作为电阻和电动势的实物基准,不过它们的制造工艺、稳定性和其他性能指标都较以前有了很大的改进和提高。

近年来,一些国家已研制出一种新的由四种合金构成的标准电阻,与锰铜电阻相比,其阻值的年稳定性提高了一个数量级,温度系数和压力系数减小了一至两个数量级。随着电子技术的发展,一种新的电压标准——固体电压标准(以齐纳二极管为基础的电压标准)已得到广泛的应用。由于齐纳二极管是固体,不怕运输和翻倒,使用温度范围很宽,与标准电池比有许多优点。目前的有源齐纳二极管电压标准,输出电压的稳定性最高的已可达到每年 $10^{-6}$ 量级,可与一等标准电池相比。

(2) 1990 年 1 月 1 日起安培的保存——电单位的量子基准

为解决实物基准随时间变化的问题,人们早就开始寻求通过不同的物理效应来建立物理量的基准,并以尽可能高的准确度来实现它。在电磁计量领域,最早作为量子基准使用的是核磁共振法电流基准。但由于此法的复现性比实物基准的稳定性差或仅仅相当,因此,这种方法只用于电流的绝对测量,而没被用于日常的单位保存和复现。

1962 年英国科学家约瑟夫森(Josephson,1940— )预言并很快被验证的低温超导下的约瑟夫森效应(Josephson)及 1980 年西德科学家冯·克里青(V. Klitzing)教授发现的量子化霍尔效应(Quantum Hall Effect)为电磁计量领域以量子基准取代实物基准提供了现实性。如果约瑟夫森常数($K_J$)、冯·克里青常数($R_k$)准确已知的话,就可以分别利用它们来定义和保存伏特、安培单位量值。

**6. 电流强度的测量**

电流的测量不管是对科学研究还是实际应用都具有非常重要的作用。电流的测量方法很多,应该根据实际的需要来选择合适的测量方法,以下仅举两例。

(1) 基准量具——安培称测电流

“安培”这个基本单位的定义和绝对测量,正是以安培定律式为依据的。如图 3-5-5 所示,对于同平面的平行电流元,安培定律可用下面的标量式表示:

$$df_{12}=\frac{\mu_0}{4\pi}\frac{I_1I_2dl_1dl_2}{r_{12}^2} \quad (3\text{-}5\text{-}5)$$

令 $I_1=I_2=I$（譬如将两电路串联起来），则有

$$I^2=\frac{4\pi\times r_{12}^2 df_{12}}{\mu_0\times dl_1dl_2}=\frac{10^7 r_{12}^2 df_{12}}{dl_1dl_2} \quad (3\text{-}5\text{-}6)$$

式中 $dl_1$，$dl_2$ 分别为两电流元的长度；$r_{12}$ 为两电流元之间距离；$df_{12}$ 为电流元 2 作用于电流元 1 上的力。

在实际中，根据上述定义式测量导线之间作用力时，不可能去测两个电流元之间的作用力，而是要测出闭合电路之间的作用力。载流回路之间的相互作用力的表达式可由安培定律导出。回路的形状采用一对平行的固定线圈 A、B 和一个动线圈 C，它们之间的作用力用图 3-5-4 所示的天平来测量。这种用来测量载流导线受磁场作用力的天平叫作安培秤。

（2）伏特计法测电流

令待测电流通过一个已知电阻 $R_0$，用伏特计测出上的电压降 $U$，如图 3-5-6 为伏特计法测电流原理图，则：$I=\frac{U}{R}$，式中 $R$ 为 $R_0$ 与伏特计内阻 $R_v$ 的并联值。

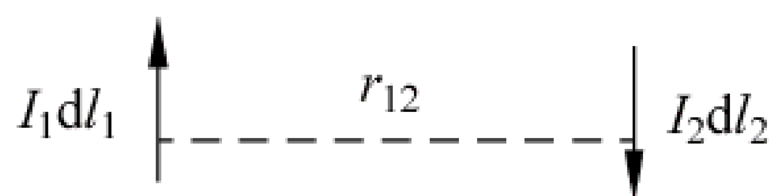


图 3-5-5　同平面的平行电流元

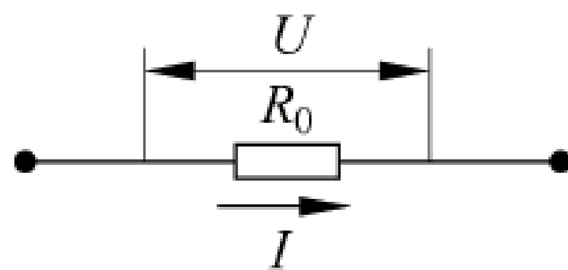


图 3-5-6　伏特计法测电流原理图

伏特计法适于用高输入阻抗的仪器测出电压 $U$，例如电子伏特计、示波器等，因为当所测电流在 $10^{-10}$A 以下时，即使用高电阻，所产生的电压降 $U$ 也很小，带来较大误差。

## 3.6　发光强度

发光强度被确定为基本物理量经历了比较复杂的过程，其主要原因在于“发光强度”是一个需要结合人类主观感觉（视觉强度）和客观描述发光体发可见光强度的物理量。1954 年，第十届国际计量大会为实现第九届国际计量大会所表示的建立一种用于国际关系的实用计量单位制的愿望，决定采用米、千克、秒、安培、热力学温度和坎德拉等 6 个单位为所要建立的这种单位制的基本单位。

1956 年，国际计量委员会决定把以第十届国际计量大会通过的基本单位为基础的单位制命名为“国际单位制”。至此，光度学单位坎德拉正式成为了国际通用的光学基本单位。

### 3.6.1　定义发光强度单位的发展历史

**1. 1860 年**

最早采用的发光强度单位是“candle”，我国旧译为“烛光”。世界上最先采用标准光源的是英国，1860 年英国首先在伦敦煤气条例上对烛光作了法律规定。顾名思义，烛光就是一支蜡烛的发光强度。显然，为保证烛光这个单位的量值不变，就得规定这种标准蜡烛的成分、尺寸、燃烧速度等，这不是一件轻而易举的事情。在此后的相当长时间内，不少科学工作者在努力研制燃烧稳定并能复制的蜡烛或油灯。1881 年，国际电工技术委员会将烛光批准为国际标准，其定义为：把 1 磅鲸脑油制成直径 7/8 英寸的 6 支蜡烛，并以每小时 120 格令

($1.43\times10^{-4}$ 磅)的速度燃烧时在水平方向上的发光强度。

1884 年,德、俄等一些欧洲大陆国家使用戊基醋酸盐为燃料的 Hefner(亥夫纳)灯,来定义发光强度的绝对单位(HK)。采用上述光源,发光强度不仅与燃料有关,而且还与灯芯火焰高度等因素有关。这些外界条件的影响,使得发光强度的复现性很差,不确定度为 6%左右。因此,当碳丝灯泡一问世,便马上被一些国家用作标准光源。白炽灯的优点在于只要灯的工作电流或电压保持不变,它的发光强度就能较准确地重现,并且使用也很方便。

**2. 1909 年**

1909 年,英、美、法等国的国家研究所决定,把各自从火焰光标准复现的单位"烛光"的平均值复现到一组碳丝白炽灯上,这组灯所体现的发光强度单位称为"国际烛光",并把保存在美国标准局的一组 45 个碳丝白炽灯的发光强度的平均值称为 1 国际烛光(International candle),符号定为 C。国际烛光这一单位于 1921 年被 CIE(国际发光照明委员会的简称,原文为法文:Commission Internationale de L'Eclairage,其英文:International Commission on Illumination,这个委员会创建的目的是要建立一套界定和测量色彩的技术标准。)承认为国际通用的发光强度单位。但德国仍坚持继续使用他们由亥夫纳灯复现的发光强度单位 HK(1HK 约等于 0.9 国际烛光)。

白炽灯作为发光强度的标准灯,同时白炽灯具有发光稳定、重复性好等优点。但是,一旦损坏就不可能再制造出第二个发光强度完全一样的灯泡。因此,它只适合于作为保存量值和传递量值的标准器,而不能作为复现单位的基准器。

**3. 1948 年**

在 1948 年 1 月 1 日,第九届国际计量大会上通过了新的发光强度单位,它的定义为:在铂的凝固点温度下,黑体发出的光亮度为 60 新烛光每平方厘米。新烛光比原来的火焰光所复现的单位约小 1.9%,单位的名称也由原来的英文"candle"改为拉丁文"candela"。我国在推行国际单位制时,单位名统一采用音译,译为"坎德拉"。

1960 年第十一届国际计量大会将坎德拉规定为基本单位之一。但因为原定义仍欠严密,所以在 1967 年第十三届国际计量大会上修改为:"坎德拉(candela)是在 101325 帕压力下,处于铂凝固点温度的黑体的 1/600000 平方米表面在垂直方向上的发光强度。"

作这一修改的原因是考虑到金属凝固点温度与其表面所受压力有关。修改后坎德拉单位的量值大小保持不变。

中、美、法、英、苏、日、加等国先后各自独立地复现了该定义。各国所用的方法和装置大体相同。

### 3.6.2 发光强度单位的最新定义

在 20 世纪六七十年代,光辐射的绝对测量技术有了很大的进展。其主要标志是出现了精密的绝对辐射计,可以用瓦特为单位,精确地测定光辐射的总功率。所以 1979 年第十六届国际计量大会通过了关于重新定义发光强度单位坎德拉的决议。决议内容为:坎德拉是发出频率为 $540\times10^{12}$ Hz 的单色辐射的光源在给定方向的发光强度,该方向上的辐射强度为 $1/683\text{Wsr}^{-1}$。

这个定义有几个新的特点:

(1) 使基本单位只和单色辐射相联系,避免了过去的复色光源基准器(如蜡烛、火焰、铂凝固点辐射)。

(2) 使基本光学单位与辐射学单位相联系,从而可以统一起来。

(3) 以频率代替波长来规定单色辐射。频率与空气折射率无关,更为严密。

(4) $540\times10^{12}$ Hz 对应于空气中 555nm 单色辐射的波长,是人眼最灵敏的区域。

(5) 定义中采用了最大光谱光视效能 $K_m$,这个值是以许多国家的实验结果为依据的。采用这一数值既保证了光度学量值的延续性,又排除了各国采用不同的 $K_m$ 值所带来的不确定性。

(6) 定义没有限定复现坎德拉的手段,有利于各国发展测量技术。

### 3.6.3 发光强度的测量及其单位“坎德拉”的复现、传递实验

**1. 光度学简介**

(1) 光度学的研究对象

光度学是光学中以人眼视觉为基础发展起来的一门研究可见光的分支学科,它发展得比较早。

辐射是以电磁波或光子的形式发射或传播的一种特殊形态的能量。辐射作用于物质或生物,就会产生各种物理、化学的或生物的效应。波长在 380~780nm 的辐射作用于眼睛,就使视觉器官产生“明”“暗”的光感觉。这是辐射的一种生物效应。光感觉的强弱不仅与辐射的功率有关,也与辐射的光谱组成有关。对同一辐射的光感觉还因人而异,并且受到心理活动的影响。例如,在黑暗的房间打开 1 千瓦的紫光灯,房间还是昏暗的;而若在同一房间中打开 1 千瓦的白炽灯时,房间就被照得很明亮了。因此,这种光感觉是一种复杂的生理-心理-物理过程。“光”这个词有时是指引起“光感觉”的物理因素,即作用于视觉器官的可见辐射(或称可见光),有时是指人体主观的“光感觉”,有时还用于泛指整个光学辐射,在不同场合有不同的含义,这是需要特别注意的。

为了适应天文学的需要,1729 年一位法国人发明了目视光度计。首先英国出现了复现发光强度单位的标准光源——特制的蜡烛。经过一百多年的发展,逐渐形成了一套测量光的体系。这个体系对光的测量完全以人的主观感觉为基础。就是说,根据这个体系所得的测量结果仅仅表示刺激视觉器官的辐射所具有的光视觉效应而不是辐射本身。

用度量光辐射强弱的物理量,如功率、功率密度等不能完全描述人眼产生的亮暗感觉。因此,在光度学研究中引入了一些与视觉有关的物理量,有发光强度、光通量、照度和亮度等。而人的视觉又是一个十分复杂的生理心理过程,因此我们常称光度学中的物理量为生理心理物理量。其中发光强度的单位坎德拉是国际单位制中的七个基本单位之一。

(2) 光度学的几个基本量

光度学的几个基本量见表 3-6-1。

**表 3-6-1 光度学中常用的物理量和单位**

| 物理量 | | 单位 | | 单位间的关系 |
|---|---|---|---|---|
| 名称 | 符号 | 名称 | 符号 | |
| 发光强度 | $I$ | 坎[德拉] | cd | |
| 光通量 | $\Phi$ | 流明 | lm | cd·sr |
| 照度 | $E$ | 勒[克斯] | lx | lm/m$^2$ |
| 亮度 | $L$ | 坎每平方米 | cd/m$^2$ | cd/m$^2$ |

(3) 视见函数

视觉器官是人体的重要器官之一。依靠视觉器官对外界刺激产生的反应,形成了明暗、

颜色、形态、远近等视知觉,以获得外部世界的信息,从而做出相应的反应。一个正常人从外界获取信息的90%以上依赖于视觉,由此可见其重要性。光度学所处理的问题直接与人类的视觉活动有关,是为人类的视觉活动服务的。人类的视觉器官的生理特性正是光度学考虑问题的出发点,是光度学的重要基础。

人类的视觉器官不仅能感受波长380～780nm范围的辐射,而且对不同波长辐射的敏感程度也不一样,这称为人眼的光谱灵敏度(人眼光谱灵敏度的另一面可看作不同波长的辐射对人眼具有不同的光视效应)。人眼的这种光谱灵敏度还受到所处环境亮度水平的影响。当亮度在几个熙提(熙提,亮度单位,简写为sb。$1sb=1lm/(cm^2 \cdot sr)$)以上时,正常人眼的适应状态叫明适应,这时的视觉叫明视觉。亮度在百分之几熙提以下时,正常人眼的适应状态叫暗适应,相应视觉叫暗视觉。处于明暗视觉之间的视觉叫介视觉。

眼睛对强光和弱光的视觉过程是由两种不同的视觉细胞来完成的,这两种感光细胞的光谱响应特性也不同。而且不同人的视觉特性也有差别。1924年国际照明委员会根据几组科学家对200多名观察者测定的结果,推荐了一个标准的明视见函数,符号为$V(\lambda)$。1933年国际计量委员会采用CIE公布的$V(\lambda)$值定义了标准光度观察者,即视见函数所代表的观察者称为CIE标准观察者。1951年CIE又公布了青年人眼的暗视觉光谱光视效率值,符号为$V'(\lambda)$。

图3-6-1给出了$V(\lambda)$和$V'(\lambda)$函数的曲线。从图中可见,视见函数曲线是一条有一中心波长,两边大致对称的光滑的钟形曲线。相应于明视觉$V(\lambda)$最大值的波长为555nm,这就是说人眼对波长为555nm的绿色光最为敏感。暗视觉$V'(\lambda)$的峰值波长为507nm。

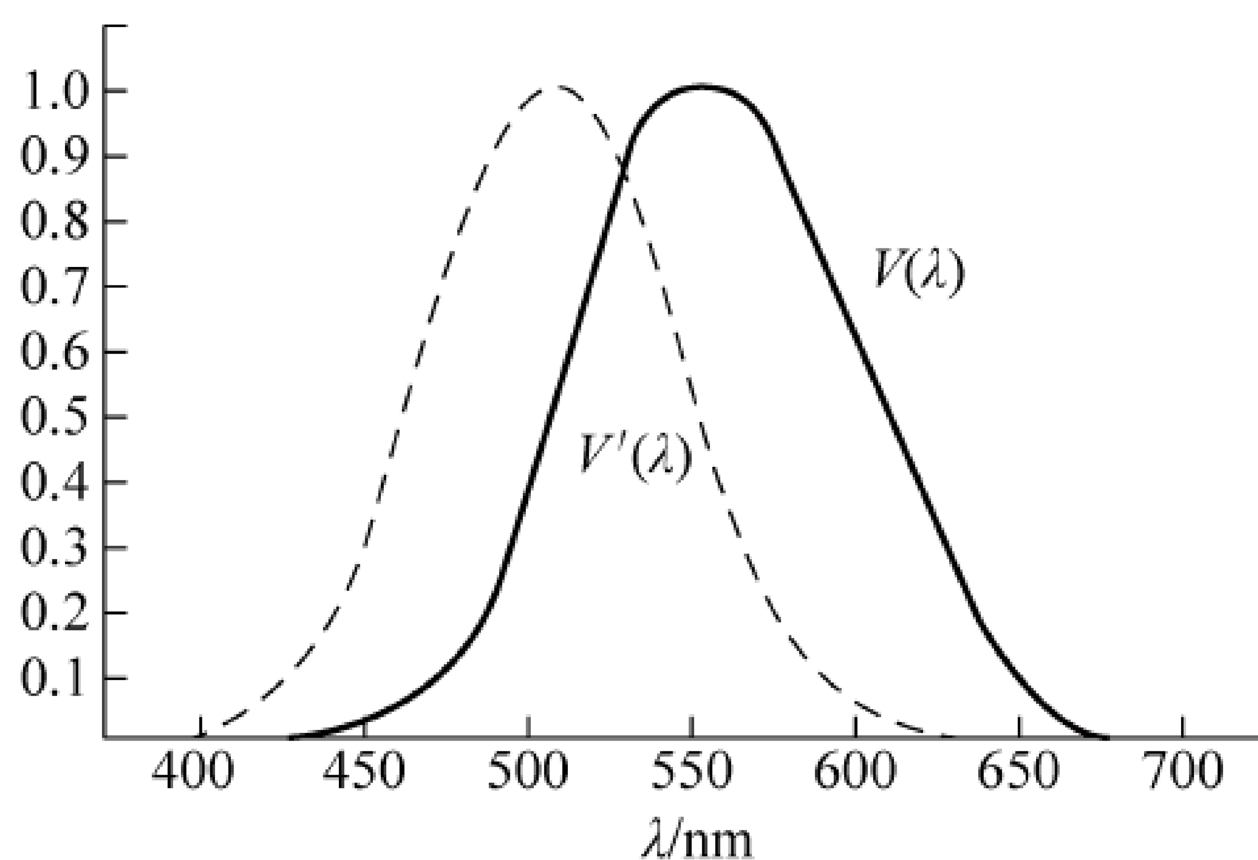


图3-6-1 明暗光谱光效率(视见函数)曲线

需要指出的是,虽然1979年以后坎德拉单位本身已和人的生理心理特性,即$V(\lambda)$无关,但这并不能改变发光强度这一物理量的性质,它仍是一个与人的生理心理因素有关的物理量。在测量光源的发光强度时,仍要用到人的视见函数$V(\lambda)$。

(4) 辐射的光视效能($K$)是辐射具有的光通量与它的功率之比,单位为lm/W。

$$K=\frac{\Phi_V}{\Phi_e} \tag{3-6-1}$$

波长为$\lambda$的单色辐射的光视效能,即光谱光视效能表示为

$$K(\lambda)=\frac{\Phi_{V,\lambda}}{\Phi_{e,\lambda}}=\frac{\mathrm{d}\Phi_{V,\lambda}/\mathrm{d}\lambda}{\mathrm{d}\Phi_{e,\lambda}/\mathrm{d}\lambda}=K_\mathrm{m}V(\lambda) \tag{3-6-2}$$

$K(\lambda)$的最大值叫作最大光谱光视效能,用符号$K_\mathrm{m}$表示。

有了上面的定义和规定,就可以将光度量和辐射度量之间的关系表示为

$$\begin{cases} X_V=K_\mathrm{m}\int X_{e,\lambda}V(\lambda)\mathrm{d}\lambda \\ X'_V=K'_\mathrm{m}\int X_{e,\lambda}V'(\lambda)\mathrm{d}\lambda \end{cases} \tag{3-6-3}$$

(5) 点光源的距离平方反比定律

假设有一个发光强度为$I$的点光源,距离点光源为$r$处有一垂直于光线方向的面元

$dA$，根据发光强度的定义，有

$$d\Phi = I d\Omega = I \frac{dA}{r^2} \tag{3-6-4}$$

式中，$d\Omega$ 为 $dA$ 面元对点光源所张的立体角。故其照度为

$$E = \frac{d\Phi}{dA} = \frac{I}{r^2} \tag{3-6-5}$$

**2. 发光强度单位的复现实验**

（1）复现原理

1979 年以后，cd 的定义改变了，1cd 定义为一个发出频率为 $540\times10^{12}$ Hz、辐射强度为 1/683$\mathrm{WSr}^{-1}$ 的单色辐射的发光强度。值得注意的是，新定义只规定了一个单色光源的发光强度值，这个光源所发出的光的频率为 $540\times10^{12}$ Hz，即在空气中的光波波长为 555nm，这个波长正好是标准明视函数的峰值波长。峰值波长所对应的光谱光视效能用 $K_m$ 来表示，单位为 lm/W。到 1977 年，包括我国在内共有 10 个国家的计量研究部门根据当时的铂凝固点黑体光度基准所体现的发光强度单位的大小，测定了 $K_m$ 值。大多数国家的测定值接近理论推算值 683 lm/W。因此在 555nm 这个波长上的 1/683W 的光功率，正好对应 1lm 的光通量。若我们注意到“强度”与“通量”这两个量的关系 $I=d\Phi/d\Omega$，就可以理解，目前坎德拉的定义，为的是保持坎德拉作为基本单位。也就是说，波长为 555nm 的单色光源的辐射强度若等于 1/683$\mathrm{Wsr}^{-1}$，正好对应其发光强度为 1 cd。

那么，对其他波长的单色光源来说，多大的辐射强度才对应于 1 cd 的发光强度呢？这就要用到 $V(\lambda)$或 $V'(\lambda)$函数了。例如对于 510nm 或 610nm 的单色光来说，由于其 $V(\lambda)$值都约等于 0.5，根据式 $K(\lambda)=K_m\cdot V(\lambda)$，故有 $K(510)=K(610)\approx 341.5$ lm/W。可见，在这两个波长上，$\frac{1}{341.5}\mathrm{Wsr}^{-1}$ 的辐射强度才对应于 1 cd 的发光强度，比峰值波长 555nm 处大了一倍。

对于复杂光谱成分的光源，例如发光稳定的白炽标准灯，其发光谱成分类似于黑体辐射的连续光谱，在实际计量这类光源时，就要用一个带有 $V(\lambda)$修正滤光器的光电接收器，使得组合后的光接收器的光谱响应度与 $V(\lambda)$函数一致。这就是目前大多数国家建立光度计量基准的主要指导思想。其主要任务是标定出一组白炽标准灯的发光强度值。

（2）发光强度复现过程

按新定义复现发光强度单位可采用电校准辐射法和光谱辐射法。因为前者比后者准确度高，国际上大多采用前者来复现坎德拉。复现 1979 年坎德拉定义的基准装置如图 3-6-2 所示。

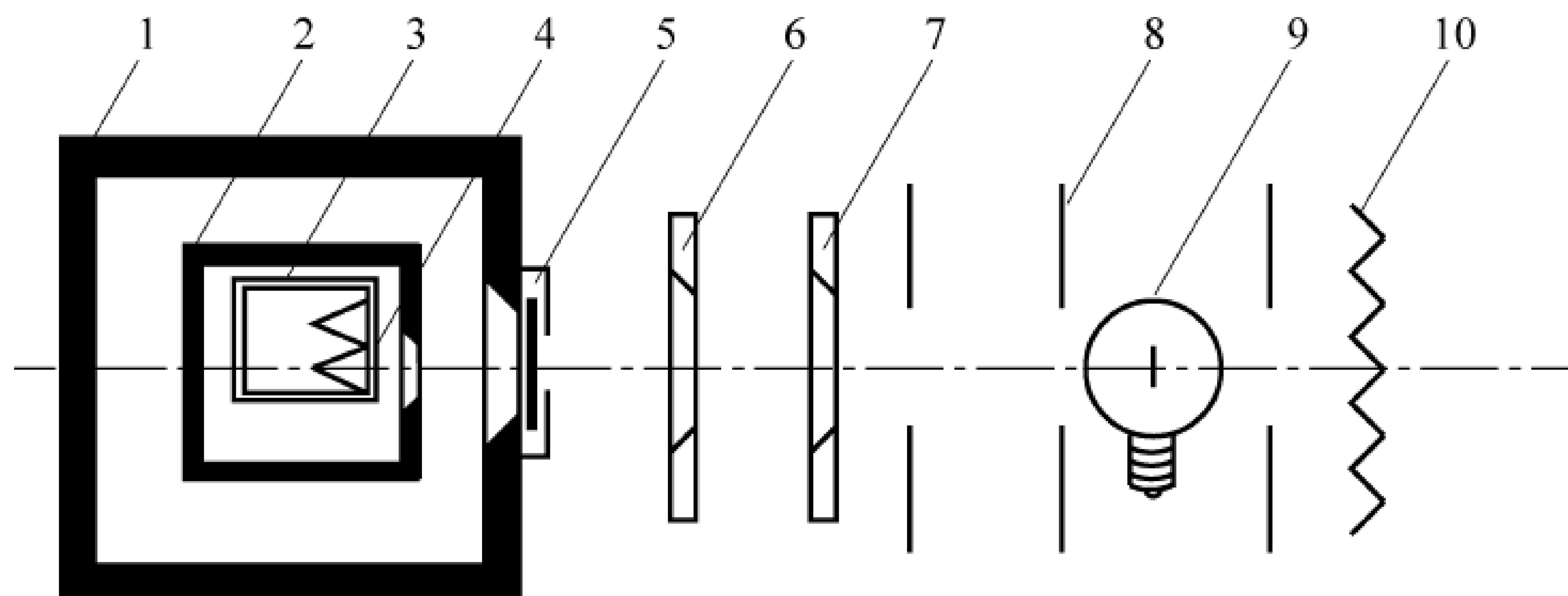



图 3-6-2　复现 1979 年坎德拉定义的基准装置

1—外屏蔽箱；2—内屏蔽箱；3—双锥腔补偿型辐射计；4—精密光阑；5—$V(\lambda)$修正滤光器及其保温套；6，7—通水挡屏；8—光阑；9—标准灯；10—光陷阱

光射到锥体内被吸收,使锥体温度升高,热电堆的热电势增大;用快门切断光源后,给加热丝通以电流,使腔体温度升高,热电堆输出的热电势达到与前者相同的值。若光加热和电加热等效,测定流过加热丝的电流和两端的电压,计算出电功率,即为待测的光功率。辐射计的功率灵敏度为 300mV/W。为了避免环境温度对辐射计的影响,把两个性能相同的锥腔型接收器极性相反地串联在一起,其中一个锥腔型接收器受光照射,另一个不受光照射,就会起补偿作用。

将锥腔型辐射计置于热屏蔽箱内,使锥腔型辐射计避免环境温度变化而产生热电势的漂移。在复现单位的实验中,辐射计是用来测定标准灯在辐射计的限制光阑表面上所产生的辐照度。$V(\lambda)$滤光器是由 4 片有色玻璃组成的,用以修正辐射计光谱响应度,使之与光谱光视效率相一致。通水挡屏是用来阻挡杂散辐射进入辐射计内的。

将灯丝平面、通水挡屏、滤光器和辐射计的限制光阑的中心精心调整在同一测量轴线上,且使灯丝平面和限制光阑垂直于该轴线,辐射计放置在热屏蔽箱内,待其温度分布均匀后再开始测量。

(3) 讨论

光源一般采用发光稳定的 BDQ-8 型光强标准灯。这个光源的发光强度可以用式(3-6-4)计算。一组约 10 只这样测定过的灯泡,构成发光强度副基准灯组,作为保存坎德拉单位的基准使用。为了测量值的高度可靠性,上述实验所用的锥腔绝对辐射计是用一组(5～7 个)结构相同,测量值一致性很好的基准绝对辐射计组成。

用上述方法得到的副基准灯组是十分宝贵的,起着保存坎德拉单位量值的作用。为减小它们的点燃时间,延长使用寿命,由副基准灯组传递到一组结构相似的工作基准灯组,作为日常工作之用。

我国于 1982 年用电校准辐射计按新定义复现发光强度单位坎德拉,其不确定度为 ±0.28%。

### 3.6.4 发光强度测量技术的展望

在实际工作中,用来体现和传递发光强度单位(cd)的量值最常用的是发光强度标准灯。作为标准光源,除了稳定性好这个最基本的要求外,还有一些性能上和结构上的要求。首先,它的使用寿命相对来讲应该较长,这里讲的寿命不是灯烧断的时候,而是它应该有一段相当长的时间发光是稳定的,这不仅与灯丝的质量有关,还要考虑泡壳的质量。也就是说,在灯泡的有效寿命之内,灯丝的结构稳定、蒸发量少、玻璃泡壳的化学性能稳定,因钨的沉积而变黑的程度轻微等。因此,作为制造标准灯的材料,钨丝应该选用优质的,并经过一定工艺处理,使其性能达到稳定。玻璃应尽量选用无色、透明度高的。在吹制泡壳时,要保证薄厚均匀,无气泡、疵点、条纹、内应力等。总体来说,新的发光材料、灯丝的排列结构、精细的做工、巧妙的设计就是标准灯一直以来在不断改进的,也是未来发展的目标。

## 3.7 摩尔

物质的量的单位——摩尔作为国际单位制的一个基本单位,是从克分子、克原子等概念发展而来的,摩尔是以物质的微观基本单元为计量对象,以阿伏伽德罗常数为计数单位的一种特殊单位。通过摩尔质量概念的建立,把微观粒子的数量跟宏观物质的质量紧密地联系起来了,如果说物质的摩尔质量是在物质的量与物质的质量之间架起的一座桥梁的话,那么

理想气体方程则是在物质的量与气态物质的体积之间架起的另一座桥梁。有了这两座桥梁，摩尔的应用便有了广阔的天地。

### 3.7.1 物质的量及其最新定义

“物质的量(amount of substance)”是国际单位制七个基本量之一，物质的量单位“摩尔(mol)”是化学计量中极其重要的单位。

**1. 摩尔的起源——从克原子、克分子到摩尔**

“摩尔”这一单位名称最初定义为：将质量以克计、数值等于分子量或原子量的物质的质量，称为摩尔。我国把它译为克分子、克原子。例如，水的分子量为18，那么18克水就是1克分子的水；硫的原子量为32，那么32克硫就是1克原子的硫。随着应用范围的拓宽，又出现了克离子、克当量、克式量等名称。

1962年以前，物理学界和化学界使用的原子量是不统一的，随之克原子、克分子也不统一。20世纪初，化学界把天然氧的原子量定为16，写为 $A_r(O)=16$。而自1929年发现氧同位素后，物理学界则更精确地把同位素$^{16}O$的原子量定为16，即 $A_r(^{16}O)=16$。天然氧是$^{16}O$、$^{17}O$和$^{18}O$的混合物。根据天然氧中上述三种核数的丰度(分别约为0.9976、0.0004和0.0020)，可以计算出氧的原子量。若以 $A_r(^{16}O)=16$，则天然氧 $A_r(O)=16.004462$，两者之比值为1.000279。科学家考察了地球上不同来源的天然氧同位素丰度的差异，将上述比值定为 $1.000275\pm0.000010$。这就是说，化学界使用的原子量比物理学界使用的原子量大了 $2.75\times10^{-4}$，由此产生的以每克分子表示的法拉第常数、气体常数和阿伏伽德罗常数也相差了 $2.75\times10^{-4}$。这样的差异对于大多数工作来说是可以忽略的，但对于精确的科学实验和热力学数据则是不允许的。这种情况延续了许多年，经化学家和物理家的反复研讨，最后于1961年取得一致意见，即规定核数$^{12}C$的原子量为12，即以$^{12}C=12$作为原子量的新标准。自此，科学界的原子量获得了统一。这时候，也只有在这个时候，为了统一化学上的克分子、克原子、克离子等概念，并为了进而更广泛地运用于物质系统的其他粒子(如电子、光子等)或这些粒子的特定组合，摩尔作为基本单位的考虑被提到了议事日程上来。此后，国际理论与应用化学联合会和国际理论与应用物理联合会共同提议，定义一个新的基本量——“物质的量(amount of substance)”，并建议以摩尔(mol)作为它的基本单位。这一建议经过国际计量委员会(CIPM)于1967年和1969年两次讨论后，于1971年第十四届CGPM通过。至此，SI的队伍中又增加了一位新的成员。

**2. 摩尔的定义及其理解**

1971年第十四届CGPM给“摩尔”规定了严格的定义：

(1)“摩尔”是一系统的物质的量，该系统中所包含的基本单元数与0.012千克$^{12}C$的原子数目相等。

(2) 在使用摩尔时，基本单位应予指明，它可以是原子、分子、离子、电子及其他微粒子，也可以是这些粒子的特定组合。

我们在理解摩尔这一概念时，上面两段话应看作是一个整体，后一段话并非可有可无。

定义中提到“与0.012千克$^{12}C$的原子数目相同”的数量究竟是多少呢？实验测定，0.012千克$^{12}C$核素所包含的碳原子数目为 $6.023\times10^{23}$。这个数字就称作阿伏伽德罗常数。由此可见，1mol某种物质所含的基本单元数目为 $6.023\times10^{23}$，故测定阿伏伽德罗常数就可以看成是对摩尔的确定。

定义中提到的“物质的量”不是指物质的质量或重量。摩尔不是克分子或克原子的同义

语。克分子或克原子的概念明确指出，它们是以克作单位的，是具有质量含义的。而物质的量则是指物质中所含微观基本单元的数量，不具有直接表示质量的含义。

摩尔既然是表示物质的量的单位，它跟一般数量单位又有什么不同呢？

这里有两点是特别值得注意的。一是它所计量的对象是微观基本单元，如分子、原子等，而不是宏观物体，不能说有几摩尔汽车、几摩尔苹果。二是它以阿伏伽德罗常数为计数单位，是个批量，而不是以1个、2个的个数来计量分子、原子等粒子的数量。这就如同在商品计量中，常以“打”为单位来计量铅笔(1打为12支)、毛巾(1打为12条)一样。

定义中所说的“基本单元”不仅仅是指分子、原子等那些构成物质的基本微粒，当然，在使用摩尔时，其基本单元经常是指分子、原子、离子等结构粒子，但它的范围并不限于此，还包括这些粒子的特定组合。例如，通常我们用摩尔来计量硫酸这种物质的量时，其基本单元就是指 $H_2SO_4$ 分子，1摩尔硫酸就含有 $6.023\times10^{23}$ 个分子，也可以取($H_2+0.234\ O_2$)，($KCl+123.4\ O_2$)作基本单元。当($H_2+0.234\ O_2$)＝1mol时，就是指该物系中含有 $6.023\times10^{23}$ 个($H_2+0.234\ O_2$)基本单元。又如，通常我们说1摩尔甲烷，指含有 $6.023\times10^{23}$ 个 $CH_4$ 分子，但在研究甲烷的键能时，则以一个“C—H”键为一个基本单元，也可以用摩尔计量。由甲烷的分子结构(图3-7-1)可以看出，一个甲烷分子中含有4个“C—H”键。这样，1摩尔 $CH_4$ 则含有4摩尔的“C—H”键。所以，通过上面的分析，我们不难看出，摩尔是以物质的微观基本单元为计量对象，以阿伏伽德罗常数为计数单位的一种特殊单位。

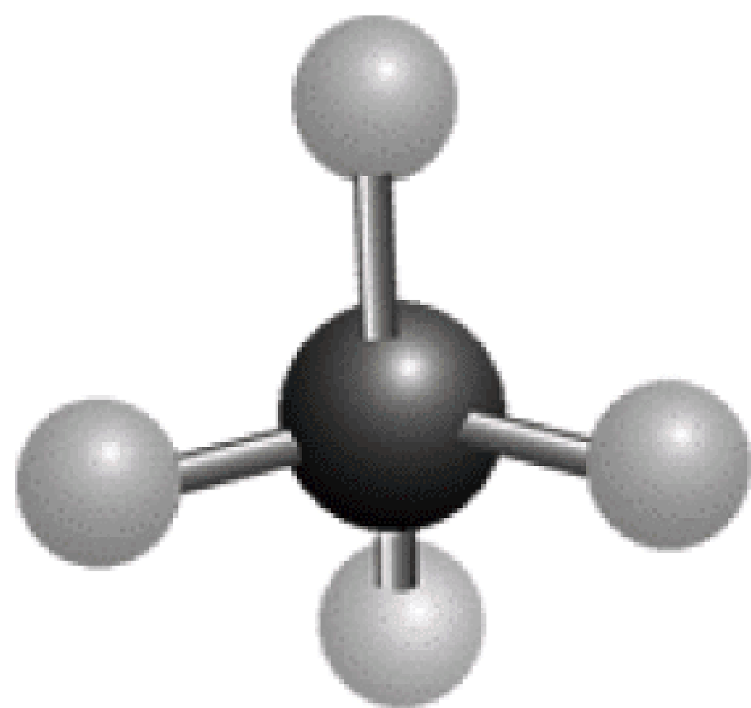


图3-7-1　甲烷的分子结构

## 3.7.2　摩尔和摩尔相关量的换算关系

在国际单位制中，物质的量单位为摩尔，符号为mol。物质的量是以它自身为量纲的量，即是说它是一个不能由其他物理量导出的一个独立的量，而且是一个基本量。

**1. 原子质量单位和相对原子质量**

为了透彻理解摩尔，必须具有关于原子量的起源和现行原子量标准的知识。原子量的现今确切说法是相对原子质量。各种元素的原子量是由国际理论与应用化学联合会(IUPAC)公布的。该联合会于1920年成立了原子量同位素丰度委员会，后者每两年要查阅和核验所有元素原子量的最新实验数据，若新数据比以前发表的数据更可靠，则要修订原先的数据。

原子量，即元素的相对原子质量，是该元素的天然核数组成的原子的平均质量对核数 $^{12}C$ 的原子质量的1/12的比值。用符号 $A_r$ 表示。它是无量纲的量。这里，$^{12}C$ 原子质量的1/12，又称为原子质量单位，其符号为u。1u等于 $1.6605402\times10^{-27}$ kg。

对任意两种元素 $x$ 和 $y$，原子质量和原子量的关系可以表示为

$$x\text{ 原子的质量}/y\text{ 原子的质量}=x\text{ 的原子量}/y\text{ 的原子量}$$

**2. 摩尔质量和摩尔浓度**

1mol物质的质量称为该物质的摩尔质量。在ISO标准中规定摩尔质量用符号 $M$ 表示。定义是：质量除以物质的量。

摩尔质量的单位是质量单位和物质的量的单位之比，是一个导出单位。质量单位是千克，物质的量单位是摩尔，所以摩尔质量单位是千克每摩尔，符号为kg/mol，通常习惯用克

每摩尔(g/mol)。建立摩尔质量的概念为我们的化学计量工作带来了很大的方便。它就像是一条纽带,把微观粒子的数量跟宏观物质的质量紧密地联系起来了,也可以说,它像一座桥梁,沟通了物质的量与物质质量间的道路。

摩尔浓度(物质的量浓度)表征单位体积的溶液中所含溶质的物质的量。物质的量单位是摩尔,体积单位是立方米,所以摩尔浓度的单位就是摩尔每立方米,其符号为 $mol/m^3$,在医学上因用 $m^3$ 太大,所以允许使用 mol/L。

**3. 气体的摩尔体积**

1mol 气体所占的体积叫作该气体的摩尔体积。在相同的温度和压强下,相同体积的任何气体,都含有相同数目的分子。特殊的,在标准状况(即 0℃,1 标准大气压)下,1mol 的任何气体,所占的体积都约为 22.4 升。摩尔体积在 ISO 标准中规定用 $V_m$ 符号表示。定义为体积除于物质的量,即

$$V_m = \frac{V}{n} \tag{3-7-1}$$

其中,$V_m$ 为摩尔体积;$V$ 为体积;$n$ 为物质的量。所以摩尔体积的单位为立方米每摩尔($m^3/mol$)。

**4. 摩尔内能、摩尔熵、摩尔热容**

摩尔内能、摩尔熵、摩尔热容的表述详见表 3-7-1。

**表 3-7-1　摩尔内能、摩尔熵、摩尔热容的表述**

| 量的名称(符号) | 量的定义 | 量 的 单 位 |
|---|---|---|
| 摩尔内能($U_m$) | $U_m = u/n$ | 焦耳每摩尔(J/mol) |
| 摩尔熵($S_m$) | $S_m = S/n$ | 焦耳每开尔文[J/(mol·K)] |
| 摩尔热容($C_m$) | $C_m = C/n$ | 焦耳每摩尔开尔文[J/(mol·K)] |

注:$u$—内能,单位为焦耳(J);
$S$—熵,单位为焦耳每开尔文(J/K);
$C$—热容,单位为焦耳每开尔文(J/K);
$n$—物质的量,单位为摩尔(mol)。

### 3.7.3　摩尔研究新进展

**1. 重新定义"摩尔"**

对阿伏伽德罗常数测量的改善可以使我们重新定义"摩尔"。根据摩尔的定义,"摩尔"是一系统的物质的量,该系统中所包含的基本单元数与 0.012 千克 $^{12}C$ 的原子数目相等,而上述基本单元数在数值上等同于阿伏伽德罗常数 $6.023\times10^{23}$,这就是摩尔与阿伏伽德罗常数的本质关系。而阿伏伽德罗常数是摩尔质量与原子质量的比值,由此根据定义应该用 0.012 千克的 $^{12}C$ 来复现摩尔,由于实践上的困难,到目前还没有人采用过 $^{12}C$ 复现摩尔。19 世纪曾经有人通过观察布朗运动估算阿伏伽德罗常数,到了 20 世纪初,科学家采用油滴实验改进了测量阿伏伽德罗常数的准确度。随着科学技术的发展,到了 20 世纪后半叶,科学家采用 X 射线衍射技术测量金属或盐的晶格间距,如测量金属钛的晶格参数来计算阿伏伽德罗常数。到了 20 世纪末,随着单晶硅制备技术的日趋完美和 X 射线及光干涉测量等技术的发展,科学家们通过单晶硅测量阿伏伽德罗常数。所以,随着测量精度的不断提高,对摩尔也必将进行重新定义。

**2. 重新定义“千克”**

众所周知,国际单位制(SI)中有7个基本单位。目前,千克在国际单位制(SI)中是唯一仍根据一种人工材料——国际千克原器而充当基准的基本单位。这实在是和科学技术日新月异的发展形势极不相称,因此人们开始研究质量“kg”的新定义。

利用原子质量或“基本”粒子(电子或质子)质量更新计量基本单位质量“kg”的定义,即用一定数量原子或“基本”粒子的质量来定义基本单位质量“kg”,这是科学家们的设想,但不论什么样的科学设想都必须确保原质量“kg”定义的延续性,即选定的(一定数量)原子或“基本”粒子的质量必须溯源在当今的千克原器上,溯源的不确定度还必须保证在原质量“kg”国际比对不确定度的范围之内($5\times10^{-8}$)。另外,为了计量基本单位传递的需要,符合质量“kg”的一定数量原子或“基本”粒子还应能构成一个宏观物体,这就是要建立宏观与微观质量单位间的关系。基本物理常量——阿伏伽德罗常数 $N_A$ 可满足这一要求。

阿伏伽德罗常数 $N_A$ 定义质量“kg”是利用下面的关系式:$1\text{kg}=\{N_A\cdot k\,\text{mol}\}\text{u}$。其中u为$^{12}$C原子质量的1/12,即原子质量单位,它由$^{12}$C同位素的质量给出;mol即摩尔,一摩尔为一系统的物质的量,该系统中所包含的基本单元数与0.012千克$^{12}$C的原子数目相等;加括号是表示括号内物理量的纯数值。所以,只要精确知道阿伏伽德罗常数 $N_A$ 的数值,就可以为千克找到以基本粒子的特性为基础的新定义。

为了得到符合要求的高纯度单晶硅、加工出符合要求的单晶硅形体和从事上述4个量准确的测量,目前国际上有4个研究单位从事该项目的科学研究,它们是:德国联邦物理技术研究院(PTB),美国国家工业标准技术研究院(NIST),意大利计量科学研究院(IMGC)和日本计量研究所(NRLM)。采用该方案测定阿伏伽德罗常数 $N_A$,测量的相对不确定度有望在五年内提高到$1\times10^{-8}$,也即与“千克”现定义的不确定度相当。

# 学习资源

## 一、中国近现代著名科学家(物理学)简介

李四光(1889—1971),字仲拱,原名李仲揆。湖北省黄冈县人,蒙古族。毕业于英国伯明翰大学,获博士学位。中央研究院院士,中国科学院院士。被尊称为“中国地质学之父”。代表作品:《地质力学之基础与方法》《地质力学概论》。李四光是世界著名的科学家、地质学家、教育家和社会活动家,是中国现代地球科学和地质工作的奠基人之一和主要领导人。中国古生物学、地质学、第四纪地质学研究的开拓者,创立大地构造理论地质力学,对亚欧大陆东部山脉体系的形成原因提出了自己独特的观点,是现代板块构造理论出现之前的大地构造理论之一。20世纪20年代,李四光在北京大学地质系担任教授。李四光对中国的能源、地震、矿产资源等方面都提出了重要的指导性意见,推动了中国地质工作的开展。他还非常精通古生物学、第四纪冰川地质和地质力学。李四光另一个重大贡献就是打破了“中国贫油论”的错误论断,为中国人民找到了大油田。1958年,李四光加入中国共产党。历任中国科学院副院长、政协全国委员会副主席、中国科学院古生物研究所所长、地质部部长、中国科学技术协会主席、中国科学院学部委员等职。曾当



选为第一、二、三届全国人大代表；第一届全国政协委员，第二、三、四届全国政协副主席；中国共产党第九届中央委员。1959 年 5 月 29 日，经苏联科学院主席团评选，授予李四光"卡尔宾斯基金质奖章"。同年 12 月 6 日，苏联地质界对他的崇高工作和在中国创建的地质科学经验，做出了公正的总结。苏联克鲁泡特金教授在《自然》杂志上撰文，介绍了李四光在地质科学上的成就。2009 年 10 月，李四光入选蒙古族十大杰出科学家。2009 年 10 月 4 日，经国际天文学联合会小天体提名委员会批准，中国科学院和国家天文台把一颗小行星命名为"李四光星"，这是继明安图后第二次用蒙古族科学家的名字命名小行星。1971 年 4 月 29 日，李四光因病逝世，享年 82 岁。

## 二、诺贝尔物理学奖简介

**1911 年诺贝尔物理学奖**。热辐射规律——维恩位移定律和建立黑体辐射的维恩公式。获奖者：维恩(德国)。评价：经过几年的安稳，经典理论的王国又开始了动摇。虽然维恩只是提出了关于黑体辐射在低频阶段的公式以及提出辐射的最大值的波长与温度的关系，并没有提出量子力学的想法，但是关于黑体辐射的研究获奖标志着此问题受到了重视，而随着此问题的进一步研究，紫外灾难的能量发散问题最终导致了量子论的发现，正如劳厄所说"他为我们打开了通往量子物理学的大门"。而瑞利和金斯两人又用经典统计力学推出了一个在高频波段和实验明显不符的结论：物体辐射在高频阶段趋于无限大。因为是从紫外频率开始所以叫紫外灾难。

**1912 年诺贝尔物理学奖**。太阳自动阀(用于白天关灯晚上开)和无人管理导航灯。获奖者：达伦(瑞典)。评价：这次的获奖没有深奥的理论和繁杂的方程，实现的理论很容易：导航灯就是高压可燃气体，而太阳阀就是利用吸收太阳的热来作为开关灯，他的获奖似乎在从另一方面诠释着物理。而他在获奖的第二年就在一次实验中炸瞎了自己的眼睛。因此他被后人称为"工程界最难得的伟大人物"。

**1913 年诺贝尔物理学奖**。物质低温性质的研究和液氦的制备。获奖者：昂尼斯(荷兰)。评价：热力学第三定律告诉我们，绝对零度不可到达但可以无限接近。昂尼斯对人类对于低温的探索做出了开创性贡献。在他的带领下莱顿大学物理实验室将当时两种还未液化的气体实现了液化，从更广的范围证明了范德瓦耳斯理论的正确性，并在研究物体低温特性时发现了超导现象，开启了物理学的一个重要新方向。

**1914 年诺贝尔物理学奖**。发现 X 射线在晶体中的衍射。获奖者：劳厄(德国)。评价：关于 X 射线的本质，自发现起就争论不休。劳厄认为 X 射线是一种电磁波，并通过晶体的衍射来证明。因为所需的狭缝太小，用机器切割绝对做不到，所以将晶体原子中的间隔作天然狭缝来形成光栅，形成了一系列斑点，称为劳厄图样。这个实验被爱因斯坦称为物理学最美的实验，证明了 X 射线的本质是电磁波。我们知道，波的另一种重要性质是衍射，其可以理解为波通过任何一个孔，产生的叠加或抵消效应。这也直接导致任何仪器都有一个分辨极限(即分辨率)。

**1915 年诺贝尔物理学奖**。X 射线研究晶体结构方面所做出的杰出贡献。获奖者 W. H 布拉格(老，英国)、W. L 布拉格(小，英国)。评价：汤姆孙的儿子，居里的女儿女婿都曾获得诺贝尔奖，而布拉格父子首创父子同时获奖的佳话。劳厄证明了 X 射线是一种电磁波，老布拉格却是此理论的反对者，绞尽脑汁设计了很多仪器想否定劳厄的结论。但小布拉格却与父亲观点相左，在父子的切磋探讨下导出了著名的布拉格公式，利用 X 射线的衍射现象来解释晶体结构。不过可惜的是英雄相轻，古今中外概莫如是。为了争谁做了主要贡献，

即使亲如父子两人还是因此失和。此外小布拉格于25岁获奖仍保持着最年轻的诺贝尔奖获得者纪录。

**1916年未颁奖**。

**1917年诺贝尔物理学奖**。发现元素的次级X射线标识谱。获奖者：巴克拉(英国)。评价：这次的颁奖依然和X射线有关,巴克拉发现通过X射线照明物体会产生一种次级射线,即只和物体本身属性有关,这标志着X射线和普通光非常相似。因为物体受到外界能量激发也会产生可见光(详见原子光谱)。不仅如此,通过这种射线的发现和布拉格公式的理论指导对研究原子内部结构有相当重要的意义。

**1918年诺贝尔物理学奖**。发现能量子(量子理论)。获奖者：普朗克(德国)。评价：关于黑体辐射问题历经了多年的研究与探索,普朗克终于解决了维恩公式和瑞利金斯公式不能适用于全频段的问题。提出了适合于全频率的普朗克公式。不过他在推导中作了一个他也无法理解的假设。光子的能量只能一份一份地传播(即只能是某个能量的整数倍),这和一般认为能量是连续的和可无限划分的思想产生了冲突。不过就是这个假设催生出了现代物理的一大基石——量子论——波粒二象性。此外普朗克曾经说过一句关于科学真理的真理,它可以叙述为"一个新的科学真理取得胜利并不是通过让它的反对者们信服并看到真理的光明,而是通过这些反对者们最终死去,熟悉它的新一代成长起来"。这一断言被称为普朗克科学定律,并广为流传。

**1919年诺贝尔物理学奖**。发现极隧射线的多普勒效应和在电场作用下光谱线的分裂现象。获奖者：斯塔克(德国)。评价：自从塞曼发现了强磁场可以使能级发生分裂后,大家就想既然磁场可以产生能级分裂那么电场也应该可以。不过包括塞曼本人在内的很多人都失败了。但斯塔克却是不费吹灰之力(他自己说的)发现了这个现象。对于此现象的解释是早期量子力学的重大胜利。

**1920年诺贝尔物理学奖**。发现镍钢合金的反常现象及其在精密物理学中的重要性。获奖者：C. E. 纪尧姆(瑞士)。评价：研究铁镍合金的过程中,发现一种含有24%的镍和2%的铬的铁合金比组成它的铁或镍具有更好的可伸展性,而对于只含有镍的铁镍合金,如果在合金中多加一些镍,那么这种合金的伸展性比起组成它的金属要差。在对镍铁合金进行了系统的研究后,他获得了一种优质镍铁合金,并把它称为铟钢,这种合金含有36%的镍,它主要的特征是在加热时膨胀系数很小。这个发明应用非常广泛,包括用来作精密测量的F-P标准具就有使用。

## 思考与练习

1. 谈谈长度单位测量技术。
2. 从物理学"质量"概念的演变,谈谈物理学理论发展的规律。
3. 时间的最新定义及其方法。
4. 为什么选取"电流"作为基本单位?
5. 试从经典物理学的观点和现代物理学的观点出发简单论述"温度"的物理意义。
6. 谈谈"负绝对温度"的本质、特点及其应用实例。
7. 详细说明"发光强度"的定义方法、物理意义和测量方法。
8. 详细说明"摩尔"的定义方法、物理意义,举例说明摩尔的测量方法。

第4章

# 大学物理学的基本理论体系

物理学曾经历了五次大型的综合，即牛顿力学的建立，热力学和统计物理学的建立，麦克斯韦电磁场理论的建立，相对论的建立，量子力学的建立。现在，物理学家正在酝酿更大范围内的综合，即统一场论的研究。

科学意义的物理学始于17世纪后半叶。当时德国的天文学家开普勒继承前人的观测资料，揭示了行星运动的三大定律。意大利科学家伽利略运用逻辑推理和精确实验推翻了亚里士多德落体运动的论点，还提出了"惯性原理"。英国科学家牛顿在此基础上推导出万有引力定律，并且在研究力学计算问题的过程中发明了微积分，在天体物体运行计算上取得成功。所以，17世纪是牛顿在伽利略、开普勒工作的基础上建立了完整的经典力学理论。此后，通过认识物体运动的本质，找到了各种现象的公共度量——能量，继而以热力学第一定律推出能量守恒定律，揭示了自然界各种运动状态的普遍性和统一性。到18世纪卡诺、焦耳等建立了宏观热力学理论；1820年奥斯特实验显示电流可以产生磁场，说明电与磁是相互联系的，法拉第运用逆向思维和对称分析法通过实验在1831年发现电磁感应定律，麦克斯韦运用他的数学才能于1868年集电磁研究之大成，把原来互相独立的电、磁和光学三个部分结合起来，创立了完整的经典电磁理论，完成了物理学第三次重大的综合和升华，形成了经典物理学宏伟大厦完善的框架。当时似乎人类对自然的认识已达到完美的境地，但是，后来的黑体辐射、光电效应、放射性的发现、电子的发现、原子的线状光谱及原子的稳定性，以及寻找"以太"实验的"零结果"等实验现象对这个"宏伟完善大厦"提出了严峻的挑战。1900年以后，经过普朗克、爱因斯坦、玻尔、卢瑟福、薛定谔和海森伯等的一大批科学家的努力，终于建立了相对论和量子力学，完成了物理学第四次和第五次重大的综合和升华。

本章将简要描述物理学各分支学科的起源、建立和发展的过程，其理论定律、原理、概念之间的相互关系。

## 4.1 力学的理论体系

物理科学的建立是从力学开始的。在物理科学中，人们曾用纯粹力学理论解释机械运动以外的各种形式的运动，如热、电磁、光、分子和原子内的运动等。当物理学摆脱了这种机械(力学)的自然观而获得健康发展时，力学则在工程技术的推动下按自身逻辑进一步演化，逐渐从物理学中独立出来。

力学又称经典力学，是研究通常尺寸的物体在受力下的形变以及速度远低于光速的情况下运动过程的一门自然科学。力学是物理学、天文学和许多工程学的基础，机械、建筑、航天器和船舰等的合理设计都必须以经典力学为基本依据。

机械运动是物质运动的最基本的形式。机械运动亦即力学运动，是物质在时间、空间中

的位置变化,包括移动、转动、流动、变形、振动、波动、扩散等。

力是物质间的一种相互作用,机械运动状态的变化是由这种相互作用引起的。静止和运动状态不变,则意味着各作用力在某种意义上的平衡。因此,力学可以说是力和机械运动的科学。

### 4.1.1 力学的起源

力学知识最早起源于对自然现象的观察和在生产劳动中的经验(人们在建筑、灌溉等劳动中使用杠杆、斜面、汲水等器具,逐渐积累起对平衡物体受力情况的认识)。古希腊的阿基米德对杠杆平衡、物体重心位置、物体在水中受到的浮力等作了系统研究,确定它们的基本规律,初步奠定了静力学即平衡理论的基础。力学的发展历程大致可分为萌芽时期、静力学发展时期、动力学发展时期、刚体力学和应用力学发展时期、近代力学发展时期和现代力学发展时期。

古代人还从对日、月运行的观察和弓箭、车轮等的使用中,了解一些简单的运动规律,如匀速的移动和转动。但是对力和运动之间的关系,只是在欧洲文艺复兴时期以后才逐渐有了正确的认识。

亚里士多德在《物理学》和《力学》两部书中对这一时期的力学知识做了概括和总结。

静力学发展时期约为公元前 3 世纪到公元 16 世纪。人们在生产劳动中应用了杠杆、螺旋、滑轮和斜面等简单机械,并逐渐总结出了力学的概念和公理,如古希腊时代就已形成密度和重心的概念,出现了杠杆原理,阿基米德于公元前 200 多年提出了浮力原理。

16 世纪以后,由于航海、战争和工业生产的需要,力学的研究得到了真正的发展,力学进入了动力学发展时期。动力学的基础由伽利略奠定。经开普勒对行星运动规律的研究,惠更斯对物理摆和曲线运动的研究,马里沃特和胡克(Robert Hooke,1635—1703)等对变形体力学的研究,至牛顿集大成而建立了以他名字命名的三大力学定律,使力学学科的科学体系得以完善。

16 世纪到 18 世纪中期是力学发展较快的时期。牛顿建立的力学体系经过伯努利(Daniel Bernoulli,1700—1782)、拉格朗日(Joseph-Louis Lagrange,1735—1813)、达朗贝尔(D'Alembert Jean Le Rond,1717—1783)等人的推广和完善,形成系统的理论,取得了广泛的应用并发展出了流体力学、弹性力学和分析力学等分支。到了 18 世纪,经典力学已经相当成熟,成了自然科学中的主导和领先学科。

此后,力学的研究对象由单个的自由质点,转向受约束的质点和受约束的质点系。这方面的标志是达朗贝尔提出的达朗贝尔原理以及拉格朗日建立的分析力学。其后,欧拉又进一步把牛顿运动定律用于刚体和理想流体的运动方程,这看作是连续介质力学的开端。

运动定律和物性方程这两者的结合,促使弹性固体力学基本理论和黏性流体力学基本理论孪生于世,在这方面做出贡献的是法国科学家纳维(Navier,1785—1836)、柯西(Cauchy. Augustin Louis,1789—1857)、泊松(Poisson. Simeon-Denis,1781—1840)、斯托克斯(George Gabriel Stokes,1819—1903)等人。弹性力学和流体力学基本方程的建立,使得力学逐渐脱离物理学而成为独立学科。

从牛顿力学到 19 世纪英国数学家哈密顿(Hamilton. William Rowan,1805—1865)的理论体系组成了物理学中的经典力学。在弹性和流体基本方程建立后,所给出的方程一时难于求解,工程技术中许多应用力学问题还须依靠经验或半经验的方法解决。这使得 19 世纪后半叶,在材料力学、结构力学同弹性力学之间,水力学和水动力学之间一直存在着风格

上的显著差别。

20 世纪初,随着新的数学理论和方法的出现,力学研究又蓬勃发展起来,创立了许多新的理论,同时也解决了工程技术中大量的关键性问题,如航空工程中的声障问题和航天工程中的热障问题等。这时的先导者是普朗特和卡门,他们在力学研究工作中善于从复杂的现象中洞察事物本质,又能寻找合适的解决问题的数学途径,逐渐形成一套特有的方法。从 20 世纪 60 年代起,计算机的应用日益广泛,力学无论在应用上或理论上都有了新的进展。

20 世纪初,相对论指出牛顿力学不适用于高速或宇宙尺度内的物体运动;20 年代,量子论指出牛顿力学不适用于微观世界。这反映人们对力学认识的深化,即认识到物质在不同层次上的机械运动规律是不同的。所以,通常理解的力学,是指以宏观的机械运动为研究内容的物理学分支学科。许多带力学名称的学科,如热力学、统计力学、相对论力学、电动力学、量子力学等,在习惯上被认为是物理学的其他分支,不属于力学的范围。

现在以牛顿定律为基础的力学理论叫牛顿力学和经典力学,它是最早形成的物理理论,曾被尊为完美而普遍理论而兴盛了 300 多年。在 20 世纪虽然发现了它的局限性,在高速领域为相对论所取代,在微观领域为量子力学所取代,但在一般技术领域包括机械制造、土木建筑,甚至航天技术中,经典力学仍保持着充沛的活力而处于基础理论的地位。

力学在中国的发展经历了一个特殊的过程。与古希腊几乎同时,中国古代对平衡和简单的运动形式就已具备相当水平的力学知识,所不同的是未建立起像阿基米德那样的理论系统。在文艺复兴前的约一千年时间内,整个欧洲的科学技术进展缓慢,而中国科学技术的综合性成果堪称卓著,其中有些在当时世界居于领先地位。这些成果反映出丰富的力学知识,但终未形成系统的力学理论。到明末清初,中国科学技术已显著落后于欧洲。

### 4.1.2 力学的学科性质

力学与数学在发展中始终相互推动,相互促进。一种力学理论往往和相应的一个数学分支相伴产生,如运动基本定律和微积分,运动方程的求解和常微分方程,弹性力学及流体力学和数学分析理论,天体力学中运动稳定性和微分方程定解理论等,因此有人甚至认为力学应该也是一门应用数学。但是力学和其他物理学分支一样,还有需要实验基础的一面,而数学寻求的是比力学更带普遍性的数学关系,两者有各自不同的研究对象。

力学不仅是一门基础科学,同时也是一门技术科学,它是许多工程技术的理论基础,又在广泛的应用过程中不断得到发展。当工程学还只分民用工程学(即土木工程学)和军事工程学两大分支时,力学在这两个分支中就已经起着举足轻重的作用。工程学越分越细,各个分支中许多关键性的进展,都有赖于力学中有关运动规律、强度、刚度等问题的解决。

力学和工程学的结合,促使了工程力学各个分支的形成和发展。现在,无论是历史较久的土木工程、建筑工程、水利工程、机械工程、船舶工程等,还是后起的航空工程、航天工程、核技术工程、生物医学工程等,都或多或少有工程力学的活动场地。

力学既是基础科学又是技术科学这种二重性,有时难免会引起分别侧重基础研究和应用研究的力学家之间的不同看法。但这种二重性也使力学家感到自豪,力学的二重性为沟通人类认识自然和改造自然两个方面做出了贡献。

### 4.1.3 力学的研究方法

力学研究方法遵循认识论的基本法则:实践——理论——实践……

力学家们根据对自然现象的观察,特别是定量观测的结果,根据生产过程中积累的经验和数据,或者根据为特定目的而设计的科学实验的结果,提炼出量与量之间的定性的或数量的关系。为了使这种关系反映事物的本质,力学家要善于抓住起主要作用的因素,屏弃或暂时屏弃一些次要因素。力学中把这种过程称为建立模型。质点、质点系、刚体、弹性固体、黏性流体、连续介质等是各种不同的模型。在模型的基础上可以运用已知的力学或物理学的规律,以及合适的数学工具,进行理论上的演绎工作,导出新的结论。

依据所得理论建立的模型是否合理,有待于新的观测、工程实践或者科学实验等加以验证。在理论演绎中,为了使理论具有更高的概括性和更广泛的适用性,往往采用一些无量纲参数,如雷诺数、马赫数、泊松比等。这些参数既反映物理本质,又是单纯的数字,不受尺寸、单位制、工程性质、实验装置类型的牵制。

因此,从局部看来,力学研究方法是多样的:有些只是纯数学的推理,甚至着眼于理论体系在逻辑上的完善化;有些着重数值方法和近似计算;有些着重实验技术等。而更多的则是着重在运用现有力学知识,解决工程技术中或探索自然界奥秘中提出的具体问题。

现代的力学实验设备,诸如大型的风洞、水洞,它们的建立和使用本身就是一个综合性的科学技术项目,需要多工种、多学科的协作。应用研究更需要对应用对象的工艺过程、材料性质、技术关键等有清楚的了解。在力学研究中既有细致的、独立的分工,又有综合的、全面的协作。

### 4.1.4 力学的学科分类

力学可粗分为静力学、运动学和动力学三部分(图 4-1-1),静力学研究力的平衡或物体的静止问题;运动学只考虑物体怎样运动,不讨论它与所受力的关系;动力学讨论物体运动与所受力的关系。

力学也可按所研究对象区分为固体力学、流体力学和一般力学三个分支(图 4-1-2),流体包括液体和气体;固体力学和流体力学可统称为连续介质力学,它们通常都采用连续介质的模型。固体力学和流体力学从力学分出后,余下的部分组成一般力学。

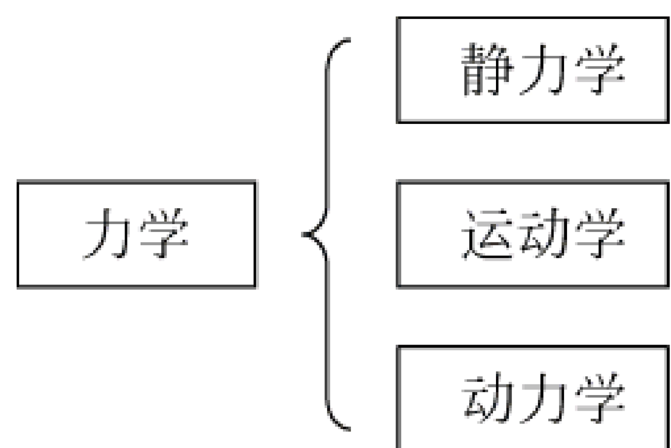


图 4-1-1 力学分类(粗分)

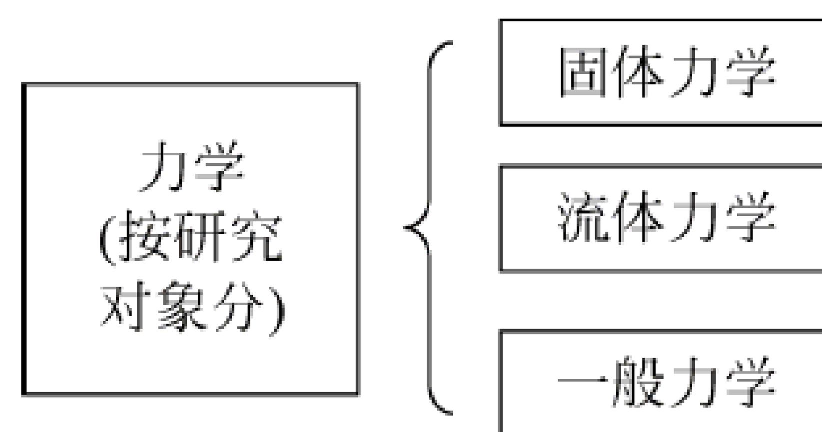


图 4-1-2 力学分类(按研究对象分)

一般力学通常是指以质点、质点系、刚体、刚体系为研究对象的力学,有时还把抽象的动力学系统也作为研究对象。一般力学除了研究离散系统的基本力学规律外,还研究某些与现代工程技术有关的新兴学科的理论。

一般力学、固体力学和流体力学这三个主要分支在发展过程中又因对象或模型的不同出现了一些分支学科和研究领域。属于一般力学的有理论力学(狭义的)、分析力学、外弹道学、振动理论、刚体动力学、陀螺力学和运动稳定性等;属于固体力学的有材料力学、结构力学、弹性力学、塑性力学和断裂力学等;流体力学是由早期水力学和水动力学这两个风格迥异的分支汇合而成,现在则有空气动力学、气体动力学、多相流体力学、渗流力学、非牛顿流体力学等分支。各分支学科间的交叉结果又产生黏弹性理论、流变学、气动弹性力学等。

力学也可按研究时所采用的主要手段区分为三个方面：理论分析、实验研究和数值计算。实验力学包括实验应力分析、水动力学实验和空气动力实验等。着重用数值计算手段的计算力学，是广泛使用电子计算机后才出现的，其中有计算结构力学、计算流体力学等。对一个具体的力学课题或研究项目，往往需要理论、实验和计算这三方面的相互配合。

力学在工程技术方面的应用结果形成工程力学或应用力学的各种分支，诸如土力学、岩石力学、爆炸力学复合材料力学、工业空气动力学、环境空气动力学等。

力学和其他基础科学的结合也产生一些交叉性的分支(图 4-1-3)，最早的是和天文学结合产生的天体力学。在 20 世纪特别是 60 年代以来，出现更多的这类交叉分支，其中有物理力学、化学流体动力学、等离子体动力学、电流体动力学、磁流体力学、热弹性力学、理性力学、生物力学、生物流变学、地质力学、地球动力学、地球构造动力学、地球流体力学等。

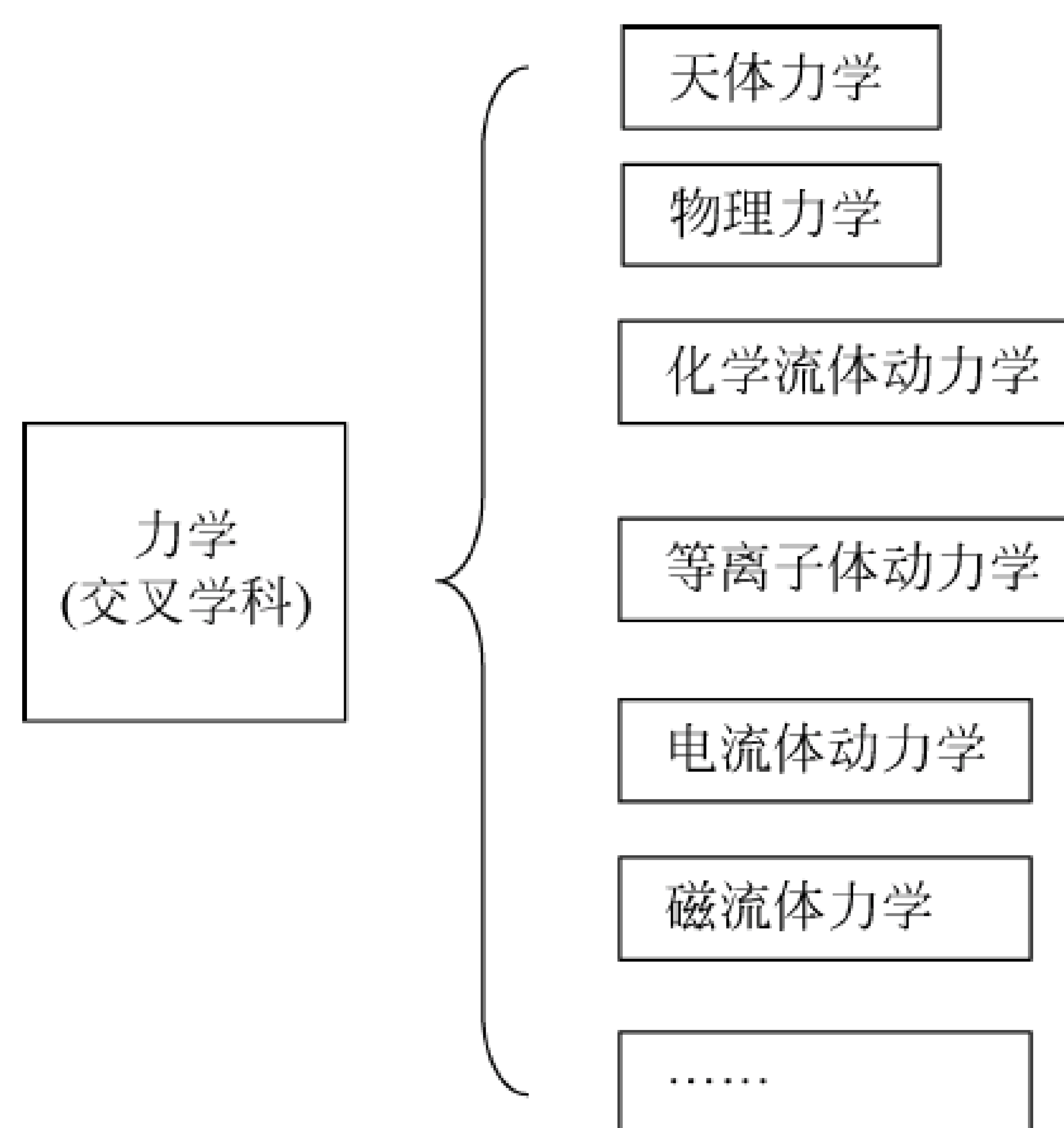



图 4-1-3　力学(交叉学科)分类

## 4.1.5　力学的基本理论

### 1. 力学的理论框架

在力学的发展史中，力学经历了经典力学、量子力学和相对论力学的阶段。在大学物理中，力学的学习内容是经典力学，它包括机械力学、流体动力学和天体力学三大内容，其理论框架如图 4-1-4 所示。

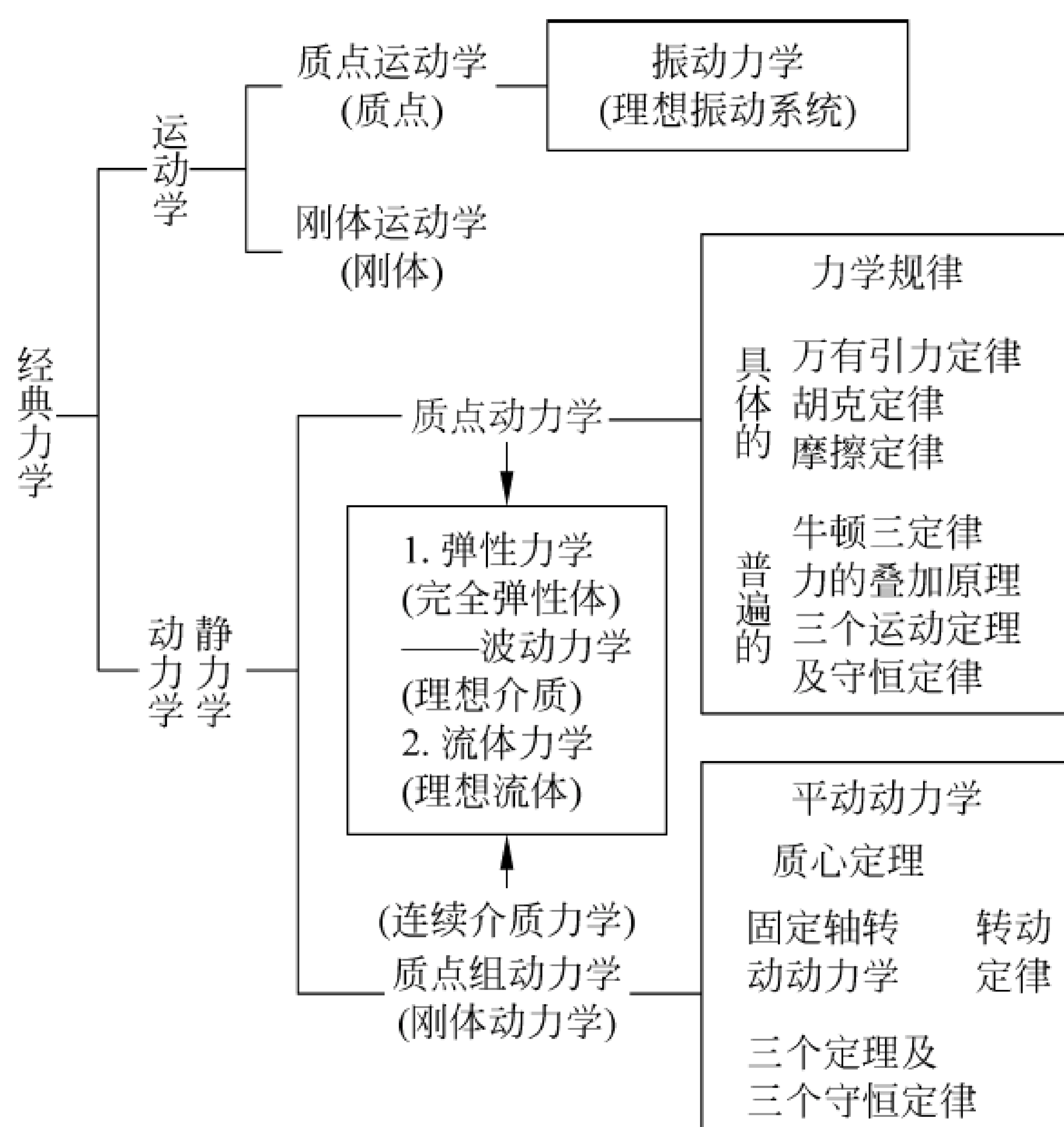



图 4-1-4　经典力学的理论框架

**2. 基本概念**

(1) 实物

力：力是物体对另一物体的作用，将受力物体视为质点，力可用受力物体对动量的变化率来量度：$\boldsymbol{F}=\dfrac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t}$。

质点：为了便于着手研究，物理学采用的方法是先做一些简化，不考虑物体的大小和形状，而把物体看作一个有质量的点，或者说用一个有质量的点来代替整个物体。用来代替物体的有质量的点叫作质点。质点是理想模型。

惯性：一切物体都有尽量维持自己原来运动状态的特性。

惯性质量：令标准物体与某物体相互作用，并用 $\Delta v_0$ 和 $\Delta v$ 分别表示标准物体与某物体速度的改变量，将与两物体有关的 $a$ 记作 $m/m_0$，有

$$m=m_0\mid\Delta v_0\mid/\mid\Delta v\mid=\mid\Delta v_0\mid/\mid\Delta v\mid\ \mathrm{kg}$$

上式所定义的为惯性质量，并简称质量。物质的多少和惯性质量目前已是不同的两种概念。

引力质量：设某质点距离地心为 $R$，受到地球的引力 $f=\dfrac{GM_{地}\,m_{引}}{R^2}$，$m_{引}$ 表示该质点的引力质量。

(2) 机械运动

惯性运动：不受力下物体静止或作匀速直线运动。

惯性参考系：孤立粒子相对它静止或作等速直线运动的参考系为惯性参考系。换句话说，在一定精度下，使牛顿第一定律成立的参考系。

非惯性参考系：使牛顿第一定律不能成立的参考系。

(3) 绝对时空观

伽利略变换：设两个惯性参考系 $O$ 系和 $O'$ 系，设时刻 $t$ 时，两坐标系的坐标原点 $O$ 与 $O'$ 重合，如图 4-1-5 所示，则某一时空，$P$ 点的坐标变换方程是

$$\begin{aligned}x'&=x-vt\\y'&=y\\z'&=z\\t'&=t\end{aligned}\tag{4-1-1}$$

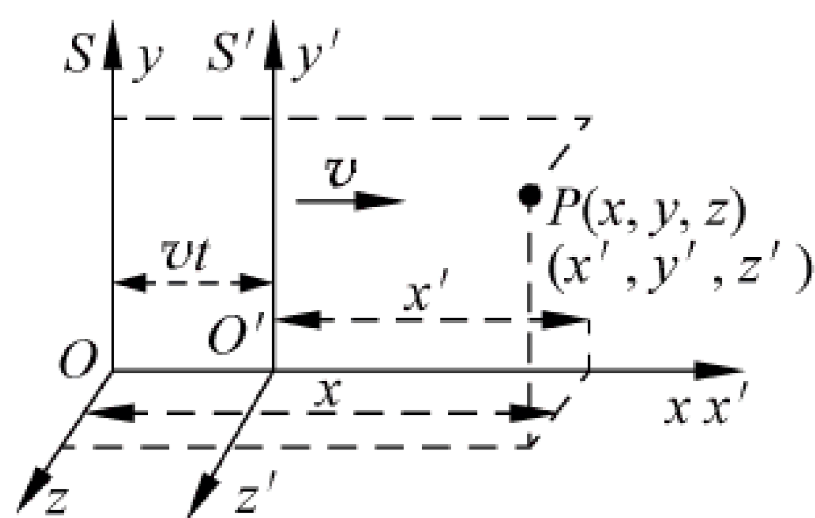


图 4-1-5 惯性参考系 $O$ 系和 $O'$ 系

这组自 $O$ 系至 $O'$ 系的时空变换关系即伽利略坐标变换方程。

这个变换方程已经对时间、空间性质作了某些假定。这些假定主要有两条：

第一，假定了时间对于一切参考系都是相同的，即假定存在着与任何具体参考系的运动状态无关的同一的时间。既然时间是不变的，那么，时间间隔在一切参考系中也都是相同的，即时间间隔与参考系的运动状态无关。时间是用钟测量的数值，这相当于假定存在不受运动状态影响的时钟。

第二，假定了在任一确定时刻，空间两点间的长度对于一切参考系都是相同的，也就是假定空间长度与任何具体参考系的运动状态无关。

牛顿说：“绝对的、真正的和数学的时间，就其本质而言，永远均匀地流逝着，与任何外界事物无关。”“绝对空间，就其本质而言，是与任何外界事物无关的，它永远不动、永远

不变。”

按照这种观点，时间和空间是彼此独立的，互不相关，并且不受物质和运动的影响。这种绝对时间可以形象地比拟为独立的不断流逝着的流水；绝对空间可比拟为能容纳宇宙万物的一个无形的、永不动的容器。

伽利略变换是绝对时空观的数学表述。

**3. 基本规律**

1）动力学规律形式不变性——伽利略变换

将 $v_{绝对}=\frac{\mathrm{d}}{\mathrm{d}t}r(t)$，$v_{牵连}=\frac{\mathrm{d}}{\mathrm{d}t}r_{o'}(t)$ 和 $v_{相对}=\frac{\mathrm{d}}{\mathrm{d}t'}r'(t')$ 分别称作绝对速度、牵连速度和相对速度。因 $t'=t$，故 $v_{相对}=\frac{\mathrm{d}}{\mathrm{d}t}r'(t')$。

伽利略变换的逆变换为 $r=r'+vt$，$t'=t$。

将逆变换对 $t$ 求导数并引入这里定义的符号，即得：$v_{绝对}=v_{相对}+v_{牵连}$。

将上式对 $t'=t$ 求导数，因 $v_{牵连}$ 为一恒矢量，故得 $a_{绝}=a_{相}$，即加速度对伽利略变换保持不变；或加速度对伽利略变换为一不变量。

2）力与物体运动的关系——牛顿第一定律、牛顿第二定律

（1）牛顿第一定律

表述1：一切物体都有尽量维持自己原来运动状态的特性，直到有外力迫使它改变为止。

表述2：孤立质点静止或作匀速直线运动。

两种表述是完全等价的。（可以讨论为什么？）

（2）牛顿第二定律

文字表述：质点的惯性质量与其加速度的乘积等于该质点所受合力，此即牛顿第二定律。

数学表达式为

$$\sum F_i = ma \tag{4-1-2}$$

3）力学规律

万有引力定律：任何两物体间均存在相互吸引力。

若物体可视作质点，则两质点的相互吸引力 $f$ 沿两质点的连线作用，与两质点的质量 $m_1$ 和 $m_2$ 成正比，与它们之间距离 $r$ 的平方成反比。数学表达式为 $f=G\frac{m_1m_2}{r^2}$，比例系数 $G$ 为对任何彼此吸引的物体都适用的普适常量，叫作引力常量，其量纲为 $L^3M^{-1}T^{-2}$。

胡克定律：对于有拉伸压缩形变的弹性体，当应变较小时，应变与应力成正比，即 $\sigma=Y\varepsilon$。

摩擦定律：

（1）摩擦力与接触面的面积大小无关，但与接触面的性质有关。

（2）摩擦力与正压力成正比，但摩擦系数不因正压力之增减而变化。

（3）动摩擦系数小于最大静摩擦系数。

（4）改变运动方向，会增大摩擦力。速度越大，则动摩擦系数越小。

（5）温度变化对摩擦力有显著影响。最大静摩擦力与正压力成正比：

$$f=\mu N \tag{4-1-3}$$

其中，$f$ 为最大静摩擦力，$N$ 为接触面之正压力，$\mu$ 称为摩擦系数。注意：正压力与接触面

垂直,摩擦力与接触面平行。

4) 力学普遍规律

牛顿第三定律:两个物体之间的作用力和反作用力总是大小相等,方向相反,作用在一条直线上,用公式表示为 $\boldsymbol{F}=-\boldsymbol{F}'$。

力的叠加原理(力的独立作用原理):如在一质点上同时作用几个力,则这些力各自产生自己的效果而不互相影响。

**4. 基本原理**

动量守恒定律:若质点系不受质点系以外其他物体的作用,该质点系动量守恒。

动能定理:质点动能的增量等于作用于质点的合力所做的功。

从微分的角度来讲,质点动能的微分等于作用于质点的合力所做的元功,即 $dA=\mathrm{d}\left(\frac{1}{2}mv^2\right)$。

功能原理:质点系机械能的增量等于一切外力和非保守内力所做功的代数和。

机械能守恒定律:在一过程中若外力不做功,又每一对非保守内力不做功,则质点系机械能守恒。

角动量守恒定律

(1) 单个质点的角动量守恒定律:若质点受一切外力对 $z$ 轴的力矩之和始终为零,则该质点对 $z$ 轴的角动量保持不变。

(2) 质点系的角动量守恒定律:几个质点分别在与 $z$ 轴垂直的平面内运动,若质点系所受一切外力对 $z$ 轴的力矩之和始终为零,则质点系对 $z$ 轴的角动量保持不变,叫作质点系对 $z$ 轴的角动量守恒定律。

**5. 惯性参考系**

惯性参考系的定义:在一定精度下,使牛顿第一定律成立的参考系。

伽利略相对性原理说明:只要找到一个惯性参考系,必然有无数个惯性参考系。

惯性参考系的选择和判断只能通过观察和试验:物体受力趋于零时可看作惯性参考系。

选择地球作为惯性参考系的条件:①时间过程短暂;②精度要求不高。

**6. 惯性**

(1) 惯性的定义:一切物体都有尽量维持自己原来运动状态的特性。

(2) 惯性的物理意义

① 牛顿第一定律中“惯性”是平动惯性,“质点”平动惯性的大小,在牛顿第二定律中惯性是质量的反映。

② 刚体的“转动惯性”用“惯性张量”表达。

③ 惯性是物体的固有属性,与外界因素无关。与该物体的“受力”或“不受力”无关。注意:惯性不是力。

④ 不能用牛顿第一定律的表述给“惯性”下定义。不能说“惯性是物体不受力下,保持静止或匀速直线运动状态的特性”。

(3) 关于惯性的进一步讨论

① 从狭义相对论的观点——“物体速度越大,其惯性越大”是对的,其依据是

$$m=\frac{m_0}{\sqrt{1-v^2/c^2}} \tag{4-1-4}$$

② 从式(4-1-4)可以看出,经典力学“物体的惯性与运动速度无关”是低速条件下的近似。

③ 如果在经典力学条件下,将物体的惯性与运动速度直接联系,则是错误的。

**7. 惯性运动**

定义:不受力时物体保持静止或作匀速直线运动称为惯性运动。

注意区分“惯性运动”与“惯性”:

(1)“惯性运动”的条件是不受力。不能将受力而合力为零条件下的运动称为惯性运动。

(2) 物体的惯性,与其“受力”或“不受力”无关。

**8. 非惯性参考系**

(1) 非惯性参考系的定义:使牛顿第一定律不能成立的参考系。

(2) 惯性力的物理意义:

由于非惯性参考系相对于惯性参考系有加速度,因而非惯性参考系中的物体由于惯性,相对于非惯性参考系有加速度,相当于物体受到一个力——惯性力,其方向与非惯性参考系相对于惯性参考系的加速度方向相反。

(3) 惯性力的类型:

(平动) 惯性力——平动非惯性参考系。

惯性离心力——转动非惯性参考系。

科里奥利力——自转动非惯性参考系。

**9. 关于牛顿第一定律和牛顿第一定律地位的讨论**

1) 牛顿第一定律是其他原理的前提和基础

(1) 首次批驳了亚里士多德——力是维持运动的原因这一错误观点;

(2) 第一次科学地给出了力的定性定义——力的本质和力的效果——(物体的)相互性,(物体的)运动状态的改变;

(3) 第一次提出了经典力学的几个基本概念——惯性参考系、力的含义(定性定义)、惯性及惯性运动;

(4) 惯性运动——回答了,如果物体“不受力”,其运动规律;

(5) 经典力学的全部规律都是建立在“惯性”参考系基础上的;

(6) 有了力的定性定义,才会有定量定义——牛顿第二定律;

(7) 有了力的完整定义,才会完善力本身的规律——牛顿第三定律、力的叠加原理、万有引力定律、胡克定律和摩擦定律。

2) 牛顿第一定律不是牛顿第二定律的特例

(1) 受力作用下物体的运动规律不能包括不受力作用下物体的运动规律,合力 $F=0$,不是不受力,虽然 $\boldsymbol{a}=\mathbf{0}$,但不是第一定律所提到的惯性运动。

(2) 牛顿第二定律是建立在物体“受力”的真实实验基础上的,因此,不能逻辑地推断出 $F\to 0$ 的理想极限情况。

(3) 牛顿第一定律引入的基本概念为牛顿第二定律的建立奠定了基础

① 惯性系概念——牛顿第二定律成立的参考系;

② 质点概念——牛顿第二定律的物质客体模型;

③ 力的定性定义——牛顿第二定律进一步定量定义力的基础;

④ 惯性概念——牛顿第二定律进一步定义惯性大小,即惯性质量的基础。

(4) 第二定律的适用范围。

在经典力学范围之外,牛顿第二定律及某些力的概念不再有效,但牛顿第一定律推广表述"一个不受外界作用的孤立体保持原来的运动状态"仍然有效。

3) 牛顿第一定律的实验验证

(1) 不能用实验直接验证

① 不受力的实验只能是理想实验;

② "惯性运动"与"不受力"是循环论证;

③ "惯性运动"与"惯性系"也是循环论证。

(2) 验证的方式是反复实验的间接论证;

(3) 逻辑推论是间接的辅助论证方法。

4) 牛顿第二定律的实验检验

(1) 在科学思维上,取决于"力"和"质量"如何定义。

(2) 在科学方法上,可以采用"控制变量法"直接验证。

**10. 经典力学理论的适用范围**

(1) "宏观"物质中的"实物","机械运动"—— 运动状态——低速运动($v \ll c$)

① "实物"——实物——质点、质点组——不包括"场";

② "机械运动"——最简单、最基本的运动形式;

③ "宏观"——经典决定论——$10^{-7} \sim 10^{5}$ m。

(2) 参考系——惯性系(弱引力场条件)

(3) 牛顿第三定律对适用范围的特殊要求

① 接触——实物接触或物体相距足够近;

② 近距离——超星系爆炸——作用力与反作用力的先后问题;

③ 低速——速度足够慢。

(4) 牛顿第二定律的适用条件

① 宏观、低速条件;

② 惯性系条件;

③ 质点条件及质点系。

关于牛顿三定律的详细讨论,请读者参阅本书5.2节。

## 4.2 热学的理论体系

热物理学是研究物质的热运动以及与热相联系的各种规律的科学。它与力学、电磁学及光学一起共同被称为经典物理学四大分支。

热物理学研究的是数量很大的微观粒子所组成的系统。这一特点决定了它有宏观的与微观的两种不同的描述方法。

这里的"微观粒子"不同于量子力学的研究对象,而是指原子、分子及其以上层次的微观颗粒。

### 4.2.1 热学的宏观描述方法与微观描述方法

热学的宏观理论是热力学,微观理论是统计物理学,两者相辅相成,构成了热学的理论

基础。

热力学是热物理学的宏观理论,它从对热现象的大量的直接观察和实验测量所总结出来的普适的基本定律出发,应用数学方法,通过逻辑推理及演绎,得出有关物质各种宏观性质之间的关系、宏观物理过程进行的方向和限度等结论。热力学基本定律是自然界中的普适规律,只要在数学推理过程中不加上其他假设,这些结论也具有同样的可靠性与普遍性。

热力学的特点:精确、可靠、普适性,但不能揭示热现象的本质。

统计物理学则是热物理学的微观描述方法,它从物质由大数分子、原子组成的前提出发,运用统计的方法,把宏观性质看作由微观粒子热运动的统计平均值所决定,由此找出微观量与宏观量之间的关系。

统计物理学的特点:近似性、普适性差;但能揭示热现象的本质。

### 4.2.2　热学发展史

18 世纪前:热现象是人类生活中最早接触到的现象之一。在有史以前人类已能钻木取火,由于火的广泛应用,人们就逐渐认识了许多热现象。但是,在古代,社会生产力很低,人们在生产和生活中对热的利用还只限于取暖和煮熟食物,最多也不过制造一些简单的金属工具。直到 18 世纪前,人类对热现象仅有粗略的了解。

18 世纪到 19 世纪:热力学形成了完整的理论体系。18 世纪初,正是资本主义发展的初期,生产中遇到的热现象大大增多,因而提供了许多关于热现象的知识。当时生产上需要动力,因而产生了利用热来获得机械功的企图。于是,就开始了对热现象进行比较广泛的研究。1714 年华侨海明改良了水银温度计,并制定了华氏温标,热学的研究走上了实验科学的道路。18 世纪中期,瓦特改进并制定成了往复式蒸汽机,利用热来获得机械功的愿望实现了。

随着蒸汽机在生产上的广泛应用,提高热机效率的问题便成了一个迫切需要解决的问题,这不仅加强了有关蒸汽机技术的研究,而且促使人们对热的本质进行深入的研究。当时关于热的本质问题,有两种观点:一是"热质说",另一观点认为热是一种运动。"热质说"曾一度在历史上占主导地位,但后来伦福德及戴维所做的实验,"热质说"无法解释,使"热质说"受到质疑,但还没有被粉碎。

1842 年德国医生迈耶主张热是一种能量并且可以与机械能互相转化,提出了能量守恒的学说。在迈耶之后的二十多年的时间里,焦耳进行了许多次实验,测定了热功当量,焦耳的实验最后确定了能量守恒与转化定律——热力学第一定律。这个定律的建立彻底粉碎了热质说。

热力学的两个基本定律都是从研究热机和机械功的相互转化问题中总结出来的。然而,热力学理论的应用远远超出了这一范围。现在,热力学理论不仅被广泛地应用在物理学各领域,而且被广泛地应用于化学生物学、医学、生态演化、天体演化、信息科学等领域。1906 年,能斯特把热力学的原理应用到低温现象和化学反应过程中,发现了一个新的规律,即绝对零度是无法达到的,这就是热学第三定律。

人们对于热的本质的认识的不断完善,也促进了分子运动论的飞跃发展。在这方面做出突出贡献的有玻耳兹曼、克劳修斯、麦克斯韦、培根、笛卡儿、胡克、罗蒙诺索夫(Михáйл Васи́льевич Ломонóсов,1711—1765)、布朗(Gordon Brown,1773—1858)等人。克劳修斯第一次清楚地说明统计的概念,正确地导出了玻意耳定律,并首先引进自由程的概念。麦克斯韦是首先认识到分子的速度各不相同而得到速度分布率的。玻耳兹曼最初在速度分布率中

引进重力场,给予熵以统计意义。这样,气体分子运动论就从一些定性的论据发展成一个系统定量的理论。

经典统计物理学的形成阶段:在1870年以后的几年里,玻耳兹曼(Ludwig Boltzmann,1844—1906)和麦克斯韦都提出了研究宏观平衡性质的概率统计法,为统计力学奠定了基石。1902年吉布斯(Josiah Willard Gibbs,1839—1903)出版了他的统计物理学书,把玻耳兹曼和麦克斯韦新创立的统计方法推广而发展成为系统的理论。

从历史上看,气体分子运动论的发展早于统计力学,可以说气体分子运动论是统计力学的前身。气体分子运动论、统计力学和涨落现象理论是统计物理学三个组成部分。

量子统计物理学的形成阶段:将统计力学应用到热辐射,发现了微观运动的新的规律性,即量子性。普朗克在1900年从热辐射的规律中发现振动的能量是能量一个最小单位的整数倍,不能连续变化,能量的这种量子化正确地解释了气体的比热容和固体的比热容随温度变化的规律,并在1926年发展成为量子力学。与比热容相应统计物理学就由经典统计物理学发展成为量子统计物理学。

### 4.2.3 热学的基本概念和原理

大学普通物理热学的基本概念和原理由三个公理及相应的一些概念组成,这些定律应用到一些具体物理过程可以得到很多重要结论,如节流过程、相变等,并构成了大学普通物理热学和热力学的完整理论体系,其内容框架如图4-2-1所示。

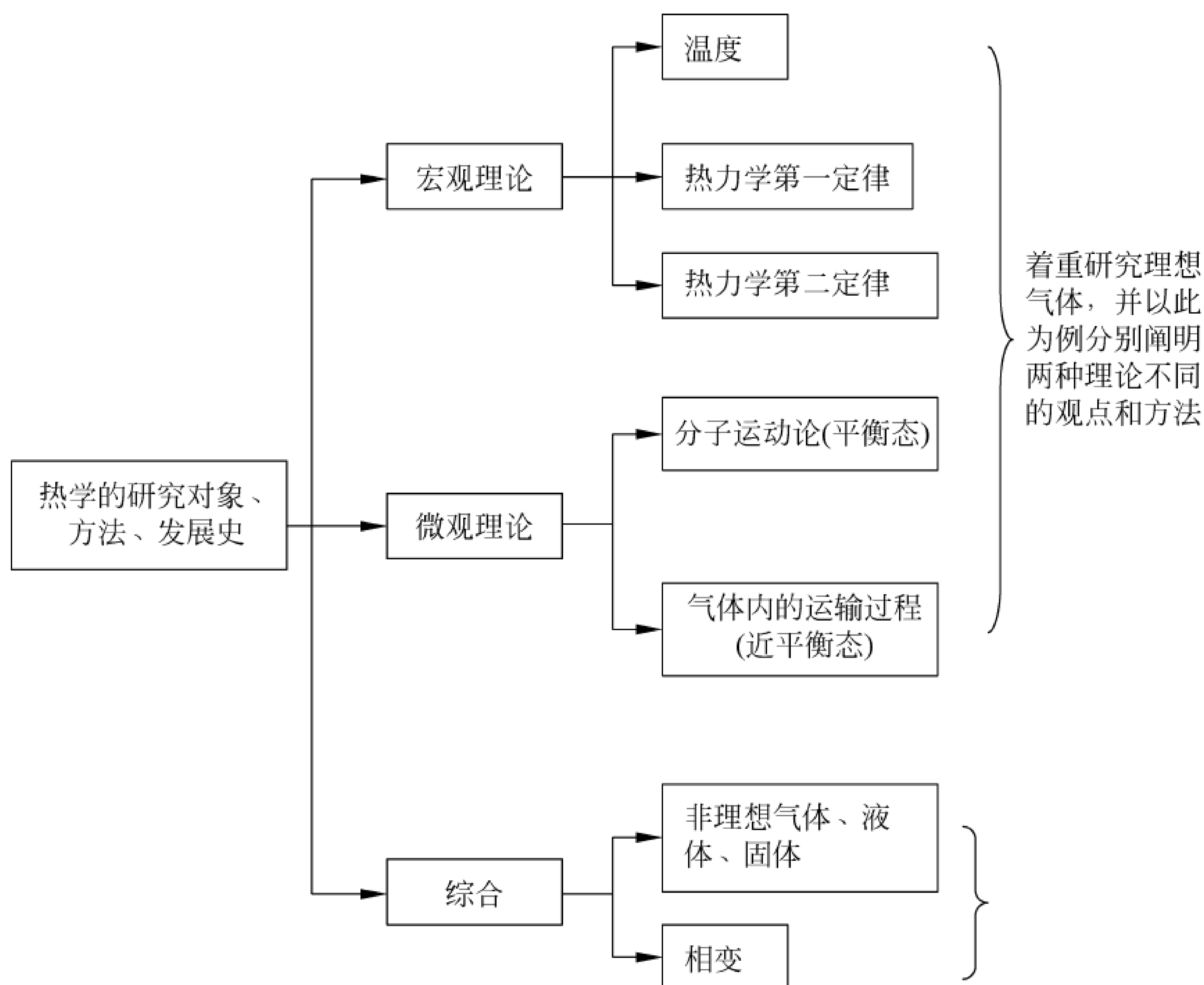



图 4-2-1 热学内容框架

**1. 温度、温标**

**定义1** 系统:物理学的任何特定分支的研究都是首先把一个限定的空间区域物质的一个有限部分从其周围分离出来,这个想象的被分离出来的进行集中研究的部分称为系统。

**定义 2**　外界：系统之外对体系的行为有直接影响的一切物质统称为外界。

**定义 3**　热力学系统：由大量分子、原子组成的物体或物体系，可用热力学变量描述的系统称为热力学系统。

**定义 4**　热力学变量：与系统内部状态有关的量、热力学系统的热性质及其热运动与机械运动等其他运动之间的转化。

**定义 5**　平衡态：不受外界影响的情况下，系统所有可观测的宏观性质都不随时间而变化的状态，不受影响指外界对系统不做功，不传热，但可受外力。

**定义 6**　热平衡：是两个（或多个）体系通过热接触后达到的状态，其特征是各系统的热力学变量取某些限定值。

**公理 1**　热力学第零定律

在不受外界影响的情况下，如果两个热力学系统中每一个都与第三个热力学系统的同一热状态处于热平衡，则这两个热力学系统彼此也必定处于热平衡。

公理 1 说明各个热力学系统都具有某种共同的宏观性质，相应地都具有某种热力学变量，它反映着各个系统热运动状态的特征，这种性质决定不同系统相互热接触时，最终必将达到热平衡，处于同一热平衡状态的所有系统，它们表征这种宏观性质的量具有相同的量值。

**定义 7**　温度：温度是决定一系统是否与其他系统处于热平衡的物理量，其数学定义式为

$$T = k \cdot E_k \tag{4-2-1}$$

其中 $E_k$ 为分子运动的平均平动能，$k$ 为比例系数。

在温度定义式 $T=kE_k$ 中，若 $k$ 为常数，对于气体则为理想气体，我们一般把很稀薄或压强很低的气体当作理想气体。

**定理**：理想气体的状态方程为

$$pV = nRT \tag{4-2-2}$$

或

$$\frac{p_1V_1}{T_1} = \frac{p_2V_2}{T_2} \tag{4-2-3}$$

其中 $R = 8.314\text{J} \cdot \text{mol}^{-1}\text{K}^{-1}$，$p_1$、$V_1$、$T_1$、$p_2$、$V_2$、$T_2$ 分别为始末状态的压强、体积、温度。

**定理**　理想气体的状态在变化过程中，其变量的变化遵守如下的过程方程：

$$pV^r = 常数 \tag{4-2-4}$$

称为多方过程，其中 $r$ 叫多方指数，如 $r=0$ 则为等压过程，$r=1$ 则为等温过程，$r=\infty$ 则为等容过程。

**定义 8**　理想气体温标：在定容条件下，规定水的三相点为 273.16K 时的压强为 $p_0$ 时，在理想气体的定容温度计的压强为 $p_v$，则

$$T_v = 273.16 p_v / p_0 \tag{4-2-5}$$

同样，可用定压理想气体温度计，测量气体的温度 $T_p$，

$$T_p = 273.16 V_p / V_0 \tag{4-2-6}$$

**2. 热力学平衡**

处于热力学平衡的系统必须满足下列严格要求：

（1）力学平衡：作用于系统的任何部分或作用于系统的整体的力都是平衡的力。

(2) 热平衡：在系统的各部分之间以及系统与环境之间不存在温度差。

(3) 化学平衡：系统内部既无化学反应，又无任何化学成分从系统的一部分向另一部分运动。

热力学系统的热力学变量随时间变化时，则认为系统经历了一个热力学过程，热力学过程由一系列中间状态组成。

**定义 9** 非静态过程：在整个热力学过程中，系统要经历一系列平衡的中间状态，这个过程叫非静态过程。

**定义 10** 准静态过程：在热力学过程进行的每一个时刻系统的状态都无限接近于平衡态。

**定义 11** 内能：热力学系统的内部状态决定的能量，它是系统状态的单值函数，在不做功的传热过程中系统内能的变化量度，称为热量。

**定义 12** 温度：温度是判定一系统是否与其他系统互为热平衡的标志。由此可以引入一个状态函数温度 $T=T(p,V,\sigma,\cdots)$。

关于温度的进一步讨论，读者可以参阅 6.2 节和 3.6 节。

**公理 2** 热力学第一定律

热力学系统的某一状态经过任意过程到达另一状态时，系统内能增量等于在这一过程中，外界对系统所做的功和系统吸收的热量的总和。

用 $U$，$Q$ 和 $A$ 分别表示系统内能，吸收的热量和外界对系统做的功，则公理 2 的数学表达式为

$$U_2-U_1=Q+A \tag{4-2-7}$$

或

$$Q=U_2-U_1-A \tag{4-2-8}$$

对于无穷小变化过程，可得公理 2 的另一数学表达式为

$$\mathrm{d}u=\delta Q+\delta A \tag{4-2-9}$$

其中 $\delta Q$，$\delta A$ 表示无穷小过程的无穷小量。

热力学第一定律反映了能量守恒和转换时应该遵从的关系，它引进了系统的态函数——内能。热力学第一定律也可以表述为：第一类永动机是不可能造成的。

**定义 13** 热容：一系统由于吸收一微小热量 $\mathrm{d}Q$ 而温度升高 $\mathrm{d}T$ 时，$\mathrm{d}Q/\mathrm{d}T$ 这个量即为热容(以 $C$ 表示)，由于热量为过程量，则热容亦为过程量：

$$C\equiv\frac{\delta Q}{\mathrm{d}T} \tag{4-2-10}$$

比热容：由于 $C\equiv\dfrac{\delta Q}{\mathrm{d}T}$ 与质量有关，则有比热容：

$$C_{\mathrm{m}}=C/m \tag{4-2-11}$$

摩尔热容：

$$C_\mu=\frac{C}{\mu} \tag{4-2-12}$$

其中 $\mu$ 为物质的量，则有

$$C_{\mathrm{m}}\mu=\frac{\mu C}{m}=\lambda C \tag{4-2-13}$$

或

$$C=\frac{\mu C_{\mathrm{m}}}{\lambda} \tag{4-2-14}$$

其中 $\lambda$ 为摩尔数。

**定义 14**　定容摩尔热容：$C_{V,\mathrm{m}}=\lim\limits_{\Delta T\to 0}\left(\frac{\Delta Q}{\Delta T}\right)=\left(\frac{\partial U}{\partial T}\right)_V$，定容过程有 $\Delta Q=\Delta U$。定压摩尔热容：$C_{p,\mathrm{m}}=\lim\limits_{\Delta T\to 0}\left(\frac{\Delta Q}{\Delta T}\right)=\left(\frac{\partial H}{\partial T}\right)_p$，定压过程有 $\Delta Q=\Delta(U-pH)$，其中 $H$ 是焓。脚标 $V$、$p$ 分别表示过程中体积、压强不变。

**定理**：理想气体的定容摩尔热容与定压摩尔热容之间的关系是

$$C_{p,\mathrm{m}}=C_{V,\mathrm{m}}+R \tag{4-2-15}$$

**3. 循环过程　熵**

**定义 15**　一系统由某一平衡态出发，经任意的一系列过程又回到原来平衡态的各个变化过程叫循环过程，若循环过程所经历的每一个分过程都是准静态过程，则叫准静态循环过程。

**定义 16**　系统对外界做功，且系统从高温物质吸收热量，向低温物质放出热量，这个循环叫正循环，反之是逆循环。

系统只与两恒温热源交换热量，无散热漏气现象，并且整个过程都是准静态地进行相应的循环叫卡诺循环。

**定义 17**　如果一个过程可以逆向进行，使系统和外界都恢复到原来的状态而不引起其他任何变化，则这种过程称为可逆过程，不满足这个条件的过程称为不可逆过程。

**定义 18**　热力学温标：定义 $Y=a\varphi$ 为热力学温标，其中 $\varphi$ 为在可逆卡诺循环中系统与热源间传递的热量，$a$ 为比例常数，或写为 $\left|\frac{Q_2}{Q_1}\right|=\frac{T_2}{T_1}$，这样规定的温标与具体物质及其特性无关，把水的三相点热力学温度规定为 273.16K，规定水的三相点温度的 1/273.16 为 1K，则热力学温标读数完全可以确定了。

**定理**：在理想气体温度计适用的范围内，理想气体温标与热力学温标相一致。

**定理**：对于一个任意的可逆卡诺循环，存在以下式子：

$$\oint\frac{\delta Q}{T}=0 \tag{4-2-16}$$

**定理**：可逆循环过程的温比热量的积分与路径无关，只与初态和终态有关，即若在两个状态 $a$、$b$ 之间沿任意路径 $L_1$、$L_2$ 的可逆过程有下式成立：

$$\int_{a(L_1)}^{b}\frac{\delta Q}{T}=\int_{a(L_2)}\frac{\delta Q}{T} \tag{4-2-17}$$

**定义 19**　由上一定理知热力学系统平衡态存在一个态函数，称为熵，用 $S$ 表示且有

$$S_b-S_a=\int_a^b\frac{\delta Q}{T} \tag{4-2-18}$$

取 $a$ 为基准态，则有

$$S_b=\int_a^b\frac{\delta Q}{T}+S_a \tag{4-2-19}$$

热学中一个重要的基本现象是趋向平衡态，这是一个不可逆过程。例如，使温度不同的

两个物体接触,最后到达平衡态,两物体便有相同的温度。但其逆过程,即具有相同温度的两个物体,不会自行回到温度不同的状态。

这说明不可逆过程的初态和终态间存在着某种物理性质上的差异,终态比初态具有某种优势。1854年克劳修斯引进一个函数来描述这两个状态的差别,1865年他将此函数定名为熵。

**公理3** 热力学第二定律

开尔文的表述:不可能从单一热源吸收热量,使之完全变为有用功而不产生其他影响。

克劳修斯的表述:不可能把热量从低温物体传到高温物体而不引起其他影响。

两种表述是等价的。它们分别揭示了功转变为热及热传递的不可逆性。

用熵的概念来表述热力学第二定律就是:在封闭系统中,热现象宏观过程总是向着熵增加的方向进行,当熵到达最大值时,系统到达平衡态。

第二定律的数学表达式:对于初态与终态均为平衡态的任意可逆和不可逆过程,有下式成立:

$$S_b - S_a \geqslant \int_a^b \frac{\delta Q}{T} \tag{4-2-20}$$

或

$$\mathrm{d}S \geqslant \frac{\delta Q}{T} \tag{4-2-21}$$

上两式中,等号适用于可逆过程,不等号适用于不可逆过程。第二定律的数学表述是对过程方向性的简明表述。

热力学基本方程:准静态过程的热力学第一定律数学表达式,$\mathrm{d}U=\mathrm{d}Q-p\mathrm{d}V$,由于在可逆过程中$\mathrm{d}Q=T\mathrm{d}S$,故第一定律可写为

$$\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V \tag{4-2-22}$$

这称为热力学的基本方程。对于理想气体,则

$$C_V\mathrm{d}T = T\mathrm{d}S - p\mathrm{d}V \tag{4-2-23}$$

所有可逆过程热力学基本上都从上面两个式子出发讨论问题。

**公理4** 热力学第三定律:绝对零度不可能达到。

关于热力学的进一步讨论,读者可以参阅本书第6章。

## 4.3 电磁学的理论体系

电磁学是研究静止电荷、运动电荷及其基本规律的一门学科,是物理学的一个重要的分支。科学地分析、把握电磁学的理论体系是学好电磁学、应用电磁学知识及学好物理学其他分支的重要方面,也是一个教师把握电磁学知识体系,组织教学,教好学生的关键。目前,国内外关于电磁学理论体系有不同的认识。大致可以归纳为两大类:一类是采用传统论述法,即从一些实验定律开始,逐步将其一般化,最后综合成麦克斯韦方程组的形式;另一类是采用公理论述法,一开始就给出麦克斯韦方程组,认定每一过程与一个实验定律相对应,在分析静态和时变的情况时,把它们作为一般化方程的特殊情况处理。电磁学的建立,经历了由实践到理论,又由理论到实践的反复过程,在这个漫长的过程中,电磁学逐步形成了独特的理论体系,对此,可概括为以下几个方面。

## 4.3.1　两个公理

**1. 电荷守恒定律**

大量的实验事实表明，电荷还有一个属性是守恒性，即在任何时刻，存在于孤立系统内部的正电荷与负电荷的代数和恒定不变，这一结论称为电荷守恒定律。电荷守恒定律是讨论起电、感应、极化、电流以及位移电流的根据，不仅在一切宏观过程中成立，也是一切微观过程普遍遵守的。

**2. 力的叠加原理**

力的叠加原理是力学中一个重要的原理，在电磁场学中仍然成立，是讨论各种形式的场合成结果的依据。

这两条原理是大量实验事实的总结，它们不仅是研究电磁学，而且又是研究其他学科的共同基础，因此我们可称其为两个公理。

## 4.3.2　三大实验定律

电磁学的主要任务是在三个实验定律(库仑定律、毕奥-萨伐尔定律和法拉第定律)的基础上，加上两条基本假说(涡旋电场、位移电流)，建立描述电磁场运动的基本方程-麦克斯韦方程组。库仑定律、毕奥-萨伐尔定律和法拉第定律是在三大实验基础上建立的三大定律，是整个电磁学理论体系建立的基础。

**1. 库仑的扭秤实验和库仑定律**

1785 年，库仑(Charlse-Augustin de Coulomb，1736—1806)在总结前人工作经验的基础上，用扭秤测量了两个带电小球之间的相互作用力，并在万有引力定律的启发下，建立了相距为 $r$ 的两个点电荷间的相互作用定律，这就是著名的库仑定律。

图 4-3-1 为扭秤实验的装置简图。当球上没有力作用时，棒处于一定的平衡位置。当轻棒上的球带电，同时把另一个带同种电荷的小球放在它附近，则会有电力作用在这个球上，球可以移动，使棒绕着悬挂点转动，直到悬线的扭力与电的作用力达到平衡时为止。因为悬线很细，很小的力作用在球上就能使棒显著地偏离其原来位置，转动的角度与力的大小成正比。通过改变两个电荷的大小和距离，观察偏转的角度，就可以得出两个电荷之间的力跟它们之间的距离的关系。

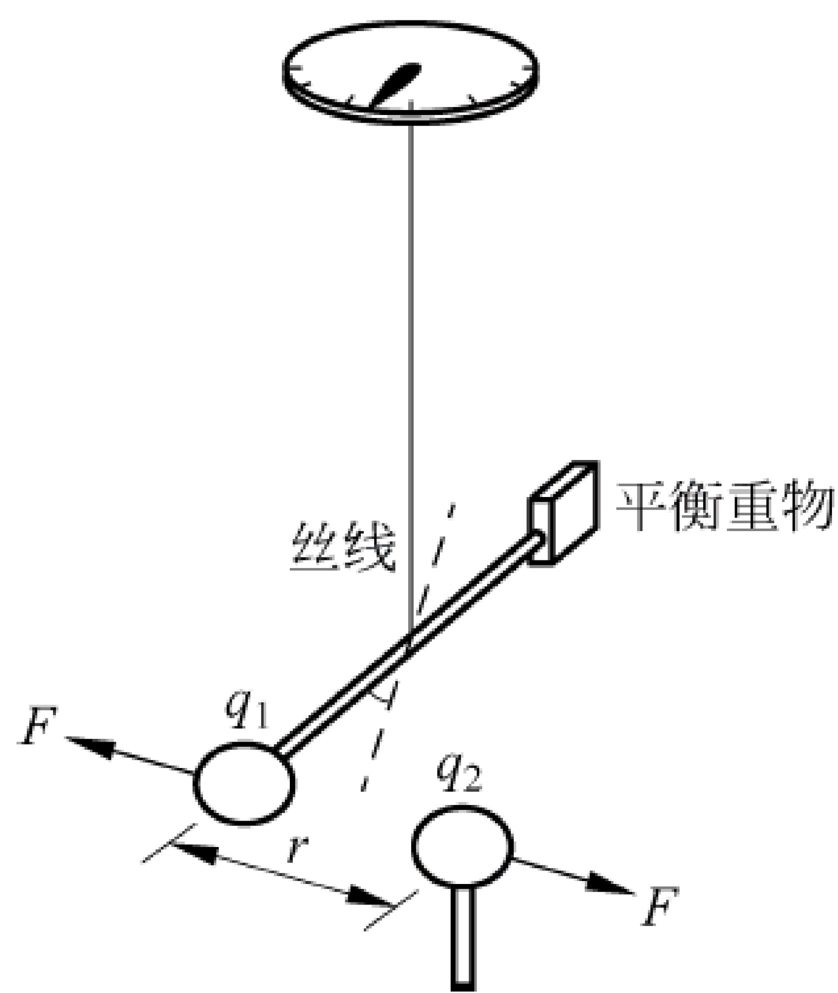


图 4-3-1　扭秤实验

库仑通过多次实验，并在万有引力定律的启发下得出两个点电荷相互作用的库仑定律：

$$\boldsymbol{F}=k\frac{q_1q_2}{r^2}\boldsymbol{e}_r \qquad (4\text{-}3\text{-}1)$$

其中，$\boldsymbol{F}$ 为两个电荷之间的作用力，$q_1$ 和 $q_2$ 分别为两个电荷的电量，$r$ 为两个电荷之间的距离，$\boldsymbol{e}_r$ 为两个电荷间的单位距离。

**2. 奥斯特电流磁效应实验和安培定律**

1820 年 7 月 21 日，丹麦科学家奥斯特发现了电流的磁效应：即在载流直导线附近，平行放置的磁针受力向垂直于导线方向偏转。奥斯特发现电流的磁效应以后，改变了把电与磁截然分开的看法，开始了探索电、磁内在联系的新时期，从此使人类认识了磁的本质，所以，奥斯特实验具有了划时代的意义。

1820年底,安培在奥斯特实验的启发下,通过实验总结出两个相距为 $r$ 的电流元 $I_1 \mathrm{d}\boldsymbol{l}_1$ 和 $I_2 \mathrm{d}\boldsymbol{l}_2$ 之间的相互作用的定量关系,又经毕奥-萨伐尔进一步实验研究,并经数学家拉普拉斯的数学加工,得出这个定量关系为

$$\mathrm{d}\boldsymbol{F}_{12}=\frac{u_0}{4\pi}\frac{I_2 \mathrm{d}\boldsymbol{l}_2\times(I_1 \mathrm{d}\boldsymbol{l}_1\times\boldsymbol{r}_{12})}{r_{12}^3} \tag{4-3-2}$$

上式为两个电流元相互作用的安培定律。

**3. 法拉第电磁感应定律**

自从奥斯特发现了电流的磁效应之后,人们自然地联想到:电流可以产生磁场,磁场是否也能产生电流呢?伟大的实验物理学家法拉第,通过大量实验终于发现,当穿过闭合导体回路中的磁通量发生变化时,回路中就出现电流,这个现象称为电磁感应现象。并于1831年总结出电磁感应定律:

$$\varepsilon_i=\frac{\mathrm{d}\varphi_{\mathrm{m}}}{\mathrm{d}t} \tag{4-3-3}$$

这个定律从现象上进一步揭示了磁与电之间的统一关系,可以称为电磁学发展史上的第三个里程碑。在此基础上,法拉第进一步提出了“场”和“力线”的概念,从此电磁学的发展进入了真正的理论阶段。

上面所讲述的三个实验定律是电磁学建立的基础,电磁学中的很多定理都可以从上面三大定律中推导得出,例如从库仑定律可以导出两条基本定理:

高斯定理:$\oiint \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S}=\frac{1}{\varepsilon_0}\sum q$ 和环路定理:$\oint \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=0$。

另外,从安培定律可以得出另外两条基本定理:

磁场中的高斯定理:$\iint \boldsymbol{B}\cdot \mathrm{d}\boldsymbol{S}=0$ 及安培环路定理:$\oint \boldsymbol{B}\cdot \mathrm{d}\boldsymbol{l}=\mu_0\sum I$ 等。

## 4.3.3 两个假设

科学发展到一定的阶段,抽象的思维和科学的假设将起关键性的作用,19世纪60年代麦克斯韦在法拉第“场”概念的启发下,科学地提出了涡旋电场假说和位移电流假说,进而建立了完整的电磁学理论体系。

**1. 涡旋电场假说**

很多实验表明,只要空间的磁场发生变化,回路中就会有感应电流。显然,该空间既无库仑力,也无洛伦兹力,究竟是什么非静电力使导体回路中的电子运动起来的呢?为了解释感生电动势的起源,麦克斯韦提出假设:随时间变化的磁场会在其周围空间激发一种电场,该电场称为感生电场,又叫涡旋电场。这样从理论上揭示了法拉第电磁感应定律的本质:变化的磁场产生涡旋电场,进而引起感应电动势,即

$$\varepsilon_i=\oint \boldsymbol{E}_i\cdot \mathrm{d}\boldsymbol{l}=-\iint \frac{\partial \boldsymbol{B}}{\partial t}\cdot \mathrm{d}\boldsymbol{S} \tag{4-3-4}$$

**2. 位移电流假说**

由于安培环路定理在非稳恒态下不成立,于是,1862年麦克斯韦由对称性出发,考虑到变化的磁场产生电场,断定变化的电场也产生一个磁场,于是提出了位移电流的假说,这个假说指出:电位移矢量对时间的变化率在磁效应方面可以等效为电流,此称为位移电流,即

$$I_d = \frac{d\phi_D}{dt} \quad 或 \quad j_d = \frac{dD}{dt} \tag{4-3-5}$$

上面 $I_d$ 称为位移电流强度，即通过电场中某截面的位移电流等于通过该截面的电位移通量 $\phi_D$ 的时间变化率；$j_d$ 称为位移电流密度，电场中某点的位移电流密度等于该点电位移矢量的时间变化率。

上述两个假设都被后来事实所证实，从而使人类对电与磁关系的认识更深了一步。

### 4.3.4 电磁学理论体系的建立

在上面三大实验定律的基础上，再加上两个假设，麦克斯韦概括出了麦克斯韦方程组，从而奠定了电磁学的理论体系。

麦克斯韦方程组如下：

$$\oint_s \boldsymbol{D} \cdot d\boldsymbol{S} = \sum q ; \oint_s \boldsymbol{E} \cdot d\boldsymbol{l} = -\frac{d\phi_m}{dt} ; \oint_s \boldsymbol{B} \cdot d\boldsymbol{S} = 0 ; \oint \boldsymbol{H} \cdot d\boldsymbol{l} = \sum I + \frac{d\phi_D}{dt}$$

必须指出的是，在介质中，上述麦克斯韦方程组尚不完备，还必须补充三个描述介质性质的方程。对于各向同性介质来说，有 $\boldsymbol{D} = \varepsilon_0 \varepsilon_r \boldsymbol{E}$；$\boldsymbol{B} = \mu_0 \mu_r \boldsymbol{H}$；$\boldsymbol{J} = \sigma \boldsymbol{E}$。

麦克斯韦方程组中，四个方程式是互相联系、互相补充、缺一不可的，其中的任何一个方程都无法凭数学手段由另外三个方程推导出来。这四个方程加在一起，全面总结了电磁学规律，利用它们原则上可以解决各种宏观电磁学问题。从麦克斯韦方程的建立过程看来，它既包括了实验概括，也包括了理论的假设和推广，它是宏观电磁学的基本方程，在电磁学中的地位和牛顿定律在力学中的地位相当。

以上的理论是电磁学的基本骨架，把握好以上理论是把握电磁学学科体系的重要方面。详细讨论请参阅本书第 7 章。

## 4.4 光学的理论体系

近代物理学的建立起源于几个典型光学实验事实与经典物理学理论的严重矛盾。可以毫不夸张地说，光学的发展史就是物理学的发展史。这不仅仅是指光学的概念和光学中的成果对物理学和物理学家的种种影响，也不只是光学仪器在物理的实验研究中占有重要的地位，更重要的是人们在研究光学的几千年的历程中，始终在追寻一个古老、现实而仍然困惑人们的基本的问题："光究竟是什么？""光速是最快的吗？"这些问题都是带有方法论意义和哲学思维本质的根本的问题。

### 4.4.1 光学的研究内容

光学的研究内容十分广泛，包括光的发射、传播和接收等规律，光与其他物质的相互作用（如光的吸收、散射和色散，光的机械作用和光的热、电、化学和生理效应等）。按照光学的发展时间，光学的研究内容可分为以下四大类：

（1）几何光学——线：直线传播、反射、折射、成像等应用光学；

（2）波动光学——波：干涉、衍射、偏振、光速——介质波；

（3）量子光学——粒子：黑体辐射、光电效应、康普顿效应、光子、德布罗意波——概率波、波粒二象性；

（4）现代光学——光子学：激光、光纤通信、非线性光学、集成光学、信息光学、超快速光学。

### 4.4.2 光学的研究方法

从方法论上看，作为物理学的一个重要学科分支，光学研究的发展也完全符合如下的认识规律：在观察和实验的基础上，对物理现象进行分析、抽象和综合，进而提出假说，形成理论，并不断反复经受实验的检验。例如围绕"光的本性是什么"这一根本问题，古往今来，人们就是沿着实验—假说—理论—实验的道路曲折前进的。这样，一方面，正确的理论对实践起指导作用；另一方面，理论通过实践又获得进一步的发展。

### 4.4.3 光学的理论体系

光是一种电磁波，电磁波满足麦克斯韦方程组及边界条件。原则上，可以由麦氏方程组及边界条件导出光的一切性质及规律。大学普通物理光学教学不是由麦氏方程组导出光的性质、规律，而是从光的几条基本原理、定律出发，得出光学的知识结构。下面从普通物理层次上对光学的理论体系进行归纳、推理，即将光学理论看成由概念、公理及由此严格导出的一系列定理所构成的理论体系。

**1. 基本概念**

**概念 1** 光：广义光指能引起效应的物质辐射能，包括可见光和不可见光。狭义光指能引起视觉感受反映客观世界形象的特殊物质，这是一个原始概念，只有把光的电磁本质揭示后，光才有一个严格的、科学的定义。

**概念 2** 波面：等相位点的集合构成的面。平面波的波面是许多相互平行的平面；球面波的波面是以光源为球心的同心球面，波前是初相位为零的波面。

**概念 3** 波线：从光源发出与波面正交指向光传播方向的曲线簇。

**概念 4** 光程：光的传播距离与其所在介质折射率的乘积的代数和。

光经过不同路线引起的光程的差值叫光程差。

**2. 几何光学**

1）三个基本实验定律

（1）光在同一均匀媒质中沿直线传播定理

如图 4-4-1 所示，自光源 $S$ 发出的球面次波在某时刻 $t$ 达到屏上圆孔 $AB$ 处，根据惠更斯原理，波面 $AB$ 上每一点都是次波源，发出球面次波，在 $t+\Delta t$ 时刻，这些次波的包络面是 $ECDF$，即为该时刻的新波面，其中 $CD$ 部分为以 $S$ 为球心，$SC$ 为半径的球面，$CD$ 之处的部分不在包络面内，所以 $SAC$，$SBD$ 为边缘光线，即光在同一介质中沿直线传播。

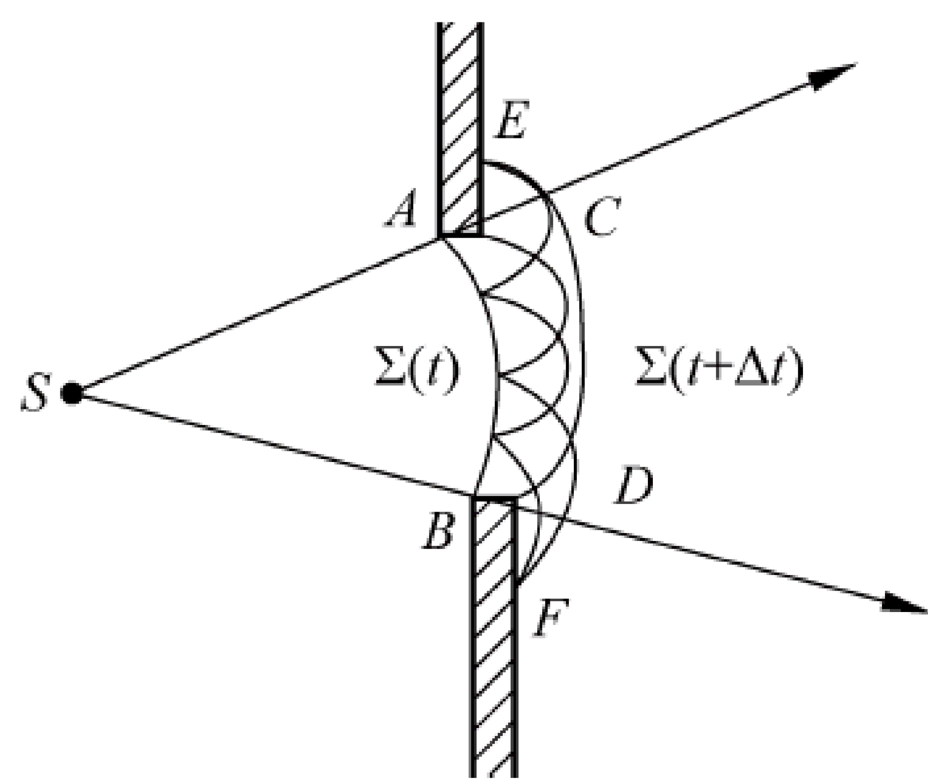


图 4-4-1 光的直线传播

（2）光通过两种介质分界面时的反射定律和折射定律

用波面、波线的概念，由图 4-4-2 可证明，$i$ 与 $i'$ 满足反射定律：$\sin i/\sin i'=\overline{DC}/\overline{AB}=v_t/v_t=1$，即

$$i=i' \tag{4-4-1}$$

反射定理可表述为：反射光线在入射面内，它与入射光线分居在法线的两侧，反射角等于入射角。

**折射定律**

用波面、波线的概念，由图 4-4-3 可证明，入射角 $i_1$ 与折射角 $i_2$ 满足折射定律：

$$\sin i/\sin r=\overline{BC}/\overline{AD}$$
$$=v_1t/v_2t=v_1/v_2=n_1/n_2 \tag{4-4-2}$$

折射定律可表述为：折射光线在入射面内，它与入射光线分居在法线两侧，折射角与入射角正弦之比与入射角大小无关，是一个与介质和光波波长有关的常数。

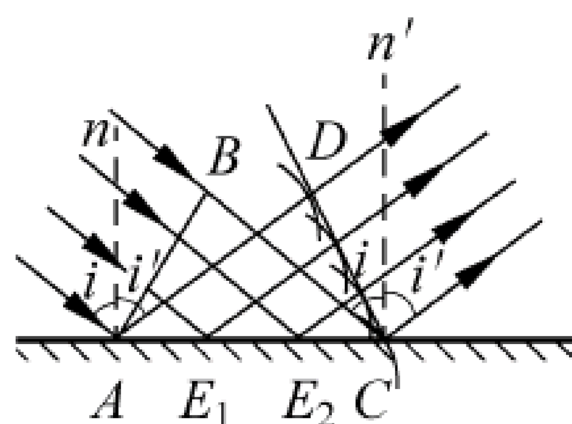


图 4-4-2　光的反射

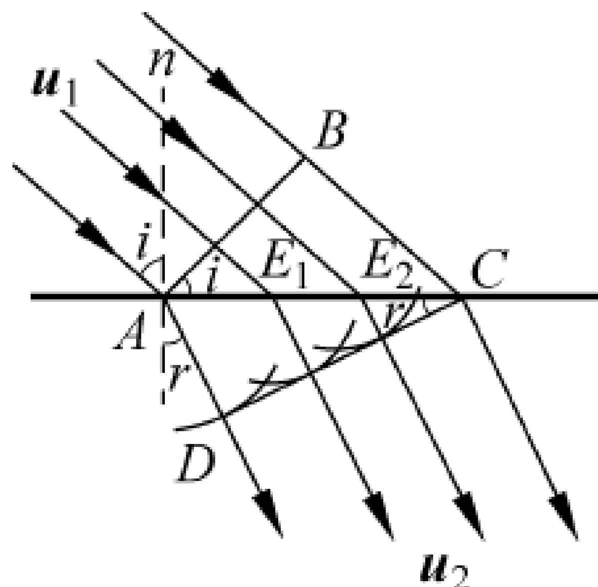


图 4-4-3　光的折射

(3) 光的独立传播定律和光路可逆原理

2) 基本原理

费马(Pierre de Fermat，1601—1665)，法国数学家、物理学家，生于博蒙德罗曼。他的职业为律师，曾任图卢兹议会的顾问 30 多年。他的一系列重要科学研究成果，都是利用业余时间完成的。在光学方面，费马确立了几何光学的重要原理，命名为费马原理。这一原理是根据经济原则提出的，是几何光学的最重要基本理论之一，对于笛卡儿的“光在密媒质中比在疏媒质中传播要快”的观点给予了有力的反驳，把几何光学的发展推向了新的阶段。

费马原理：光在指定的两点间传播，实际的光程总是一个极值。也就是说，光沿光程为最小值、最大值或恒定值的路程传播。其数学表达式如下：

$$\int_A^B n\,\mathrm{d}s=\text{极值(极小值、极大值或恒定值)} \tag{4-4-3}$$

在一般情况下，实际光程大多是取极小值，费马本人最初提出的也是最短光程。这是几何光学中的一个最普遍的基本原理，称为费马原理。

根据两点间直线距离最短这一几何公理，从费马原理可以直接推出光在均匀介质(或真空)中沿直线传播。此外，可以证明，光通过两种不同介质的分界面时，所遵从的反射定律和折射定律也是费马原理的必然结果。

3) 几何光学成像万能公式

从三个基本实验定律可推导出一个最基本的公式—— 单球面折射成像公式，又称之为成像万能公式，即阿贝不变式，其他的成像公式可从它推导出：

$$\frac{n'}{s'}-\frac{n}{s}=\frac{n'-n}{r} \tag{4-4-4}$$

式中 $n$，$n'$分别是入射、折射光线所在介质折射率；$s$，$s'$分别是物矩、像距；$r$ 是球面曲率半径。

(1) 球面镜反射成像公式

将 $i=-i'$，$n=-n'$代入式(4-4-4)，可得

$$(1/s')+(1/s)=2/r \tag{4-4-5}$$

为球面镜反射成像公式。

(2) 平面镜反射成像公式

在式(4-4-4)中，$r\to\infty$时，有 $s'=-s$，得到与物等大的虚像。

(3) 像方焦距与物方焦距

在式(4-4-3)中,若 $s\to\infty$,得像方焦距:$f'=n'r/(n'-n)$;若 $s'\to\infty$,得物方焦距,$f=-nr/(n'-n)$。

(4) 薄透镜成像高斯公式

将式(4-4-4)两边除以 $(n'-n)/r$ 可得薄透镜成像高斯公式

$$f'/s'+f/s=1 \tag{4-4-6}$$

在式(4-4-5)中,设物点到物方焦点长度为 $x$,像点到像方焦点长度为 $x'$,则得到

$$xx'=ff' \tag{4-4-7}$$

式(4-4-6)为薄透镜成像牛顿公式。

(5) 多个单球面成像公式

可依次使用式(4-4-4):

$$(ni'/si')-(ni/si)=(ni'-ni)/ri \tag{4-4-8}$$

在式(4-4-8)中令 $i=1,2,3,\cdots,m$,就得到经 $m$ 个单球面后的成像公式。对于薄透镜,两次使用式(4-4-8)得薄透镜成像公式:

$$(n_2/s_1)-(n_1/s)=(n-n_1)/r_1+(n_2-n)/r_2 \tag{4-4-9}$$

**3. 波动光学**

波动光学的基本原理:线性叠加原理和惠更斯-菲涅耳原理。

1) 线性叠加原理

(1) 原理阐述:空间某位置几列光波相遇,在该位置所引起的振动矢量等于各列波分别在该位置所引起的振动矢量的矢量和。

该原理可认为是物理学中广泛的矢量叠加原理在光学中的具体应用。

(2) 应用:光的干涉现象

两列频率相同,振动方向一致,有固定位相差的光波在波场中相遇,能量重新分布,形成明暗相间条纹的现象。光的干涉现象在理论上用光的叠加原理进行解释,并得到明暗条纹的关系式。

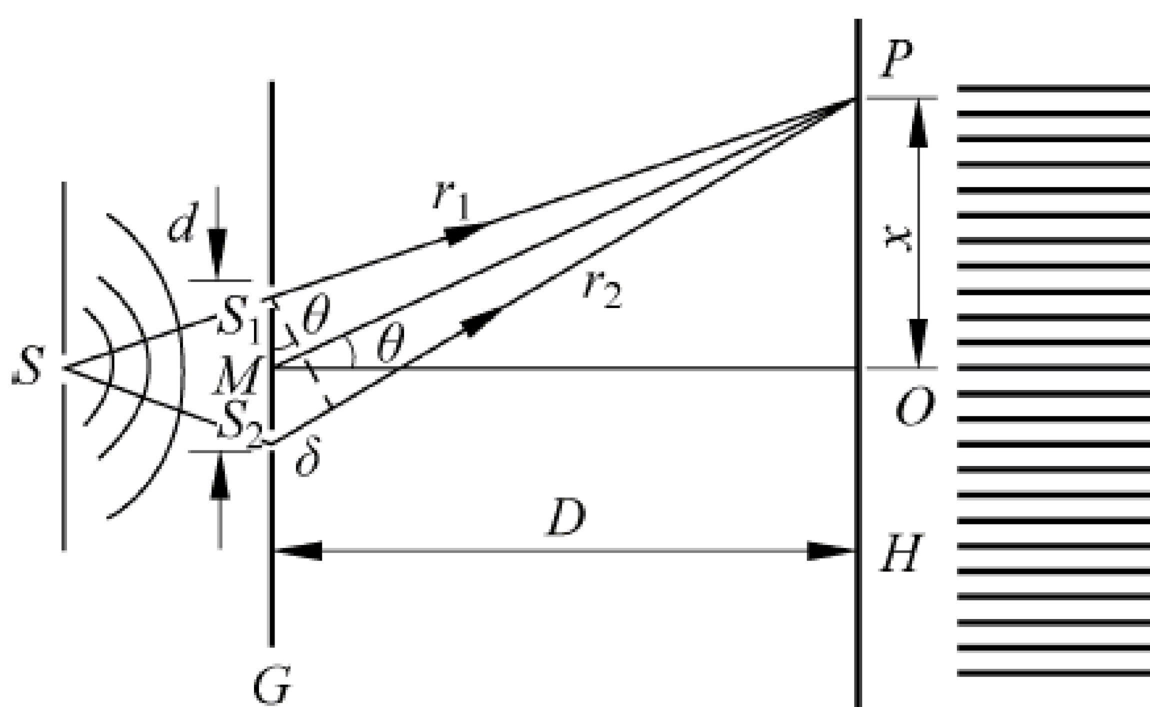



图 4-4-4　杨氏双孔干涉

杨氏双孔干涉见图 4-4-4,同一波源发出的球面次波通过 $S_1$、$S_2$ 在后面的屏上 $P$ 处相遇产生干涉。$S_1$、$S_2$ 到达 $P$ 时,振动为

$$E_1=A_1\cos[\omega t-(2\pi/\lambda)r_1] \text{ 和 } E_2=A_2\cos[\omega t-(2\pi/\lambda)r_2]$$

叠加($d\ll D$,$S_1$、$S_2$ 近似认为振动方向相同):

$$I=A_1^2+A_2^2+2A_1A_2\cos\varphi=4A_1^2\cos(\Delta\varphi/2) \tag{4-4-10}$$

其中,

$$\Delta\varphi=2\pi(r_2-r_1)/\lambda \tag{4-4-11}$$

为相位差。

① 振动加强区为相长干涉:$\Delta\varphi=2k\pi,k=0,\pm1,\pm2,\cdots$或

$$r_2-r_1=2k\,\frac{\lambda}{2},\quad k=0,\pm1,\pm2,\cdots \tag{4-4-12}$$

② 振动削弱区为相消干涉：$\Delta\varphi=(2k+1)\pi, k=0,\pm1,\pm2,\cdots$或

$$r_2-r_1=(2k+1)\frac{\lambda}{2},\quad k=0,\pm1,\pm2,\cdots \tag{4-4-13}$$

③ 结论：光程差是半波长的偶数倍时为相长干涉；光程差是半波长的奇数倍时为相消干涉。两列光波达到屏幕：

$$r_2-r_1=d\sin\theta\approx d\tan\theta=(x/D)d,\quad \Delta\varphi=(2\pi/\lambda)d(x/D)$$

由式(4-4-12)与式(4-4-13)知，当 $x=(D/d)2k(\lambda/2), k=0,\pm1,\pm2,\cdots$时 $P$ 点光强最强，为相长干涉；当 $x=(D/d)(2k+1)(\lambda/2), k=0,\pm1,\pm2,\cdots$时 $P$ 点光强最弱，为相消干涉；条纹间距 $\Delta Y=x_{k+1}-x_k=(D/d)\lambda$。

其他干涉现象如薄膜干涉、牛顿环、迈克耳孙干涉等都可以用叠加原理解释，并导出明暗条纹分布公式。

2）惠更斯-菲涅耳原理

（1）原理表述：如图 4-4-5 所示的波面 $S$ 上每个面积元 $\mathrm{d}s$ 都可以看成新的波源，它们均发出次波。波面前方空间某一点 $P$ 的振动可以由 $S$ 面上所有面积元所发出的次波在该点叠加后的合振幅来表示。

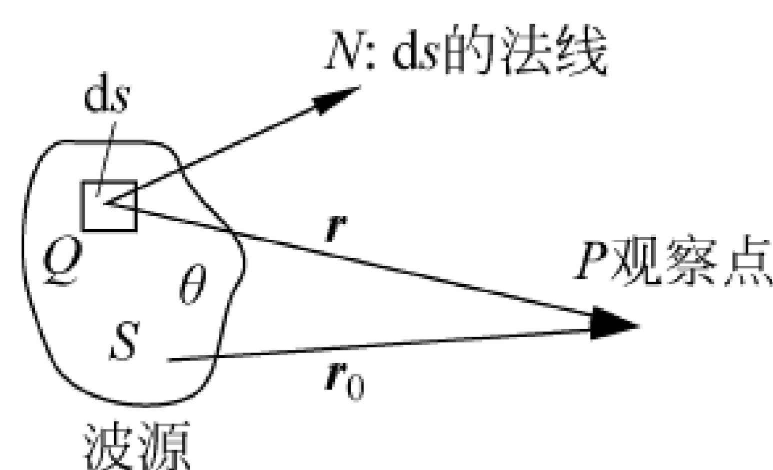



图 4-4-5　惠更斯-菲涅耳原理的计算

（2）意义及问题

① 用波动的观点解释了光的衍射现象；

② 解决了衍射光强分布的定量计算问题；

③ 复杂衍射物的数学计算非常困难。

（3）菲涅耳的衍射积分公式

在 $t_1$ 时刻，波前面元 $\mathrm{d}s$ 发出的次波，在 $P$ 点合振动 $\mathrm{d}E$（已经为 $t_2$ 时刻了）为

$$\mathrm{d}E=\frac{A(Q)k(\theta)}{r}\cos(kr-\omega t)\mathrm{d}s \tag{4-4-14}$$

其中，$\theta$ 为光源 $Q$ 点法线 $N$ 对观察 $P$ 点矢径 $\boldsymbol{r}$ 的夹角；$A(Q)$为光源 $S$ 的振幅分布函数；$k(\theta)$为光源 $Q$ 对观察点 $P$ 的倾斜因子

$$E_P=\int_S \mathrm{d}E=c\int_S\frac{A(Q)k(\theta)}{r}\cos(kr-\omega t)\mathrm{d}s$$

或

$$E_P=\int_S c\,\frac{A(Q)k(\theta)}{r}\mathrm{e}^{\mathrm{i}(kr-\omega t)}\mathrm{d}s \tag{4-4-15}$$

**注意**：只有对圆孔、圆形、矩形、方形等形状规则衍射物，才可以计算出精确结果。

（4）应用：光的衍射现象

光的衍射现象：光能绕过障碍物且（在观察屏上）出现光强不均匀分布的现象。光的衍射现象可用惠更斯-菲涅耳原理解释，并给出明暗条纹分布表达式。

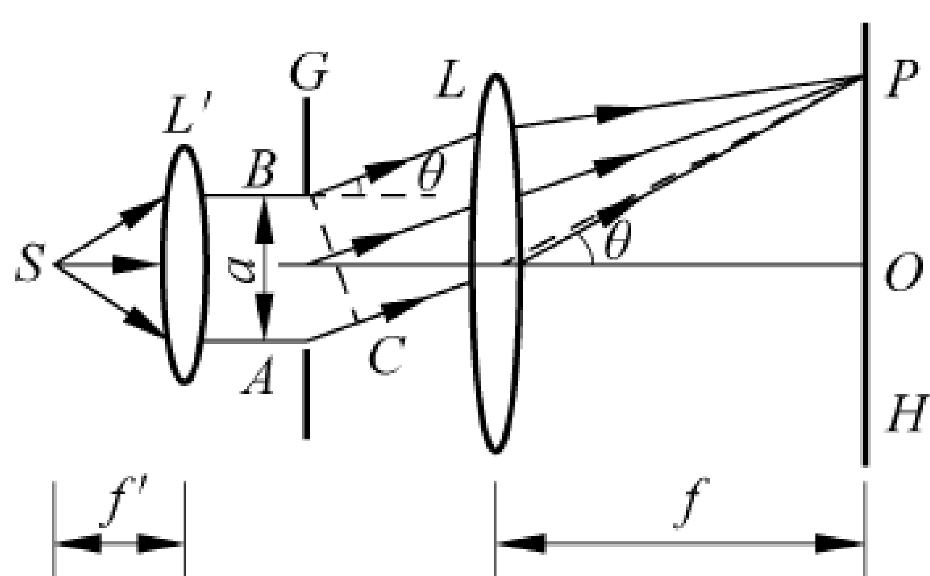



图 4-4-6　夫琅禾费单缝衍射

如图 4-4-6 为夫琅禾费单缝衍射。两条边缘光线 $a_1$，$a_2$ 在 $P$ 点路程差为 $BC=\sin\theta$，其中 $b$ 为狭缝宽度，$\theta$ 为衍射角，如光程差为 $\Delta=(2k+1)\pi/2, k=0,\pm1,\pm2,\cdots$，则 $P$ 点衍射强度为极大值，如光程差为 $\Delta=2k\cdot\pi/2, k=0,\pm1,\pm2,\cdots$，则 $P$ 点衍射强度为极小值。

其他衍射现象如圆孔衍射、平行光栅衍射都可用惠更斯-菲涅耳原理解释，并可用同样的方法导出明暗条纹分

布公式。

3）光的偏振

光是电磁波，电磁波是横波，光波也有偏振现象，光的偏振现象最明显的是反射起偏与折射起偏。在菲涅耳公式中，当 $i_0+i''=\pi/2, E'_{/\!/}=0$ 时，$\sin i_0/\sin i''=n_2/n_1$，$\sin i_0=(n_2/n_1)\sin i''=(n_2/n_1)\sin[(P/2)-i_0]=(n_2/n_1)\cos i_0$，$\tan i_0=n_{21}$，$i_0=\arctan n_{21}$。式中 $i_0$ 为入射角，$i''$ 为折射角，$E'_{/\!/}$为反射振动矢量平行于入射面的量，$n_1$、$n_2$ 为入射、折射介质的折射率。上式即为布儒斯特定律：光照到透明物质上，入射角为某一数值 $i_0$，反射光线与折射光线垂直，反射光线为完全线偏振光，其振动面垂直于入射面(在此，该定律的地位相当于一个定理)。

由光在晶体中的双折射现象也可以得到偏振光，如尼科尔棱镜等，这种现象可用惠更斯原理得到解释。惠更斯假设光在晶体中分别以球面波和椭球面波的方式传播两束光，它们都是偏振光。

**4. 量子光学**

1）黑体辐射

(1) 绝对黑体：理想模型，其吸收系数 $A(\nu,T)=1$，用空腔的小孔模拟。

(2) 绝对黑体单色辐出度——恰等于普适函数，即

$$M(\nu,T)=\frac{M(\nu,T)}{1}=\frac{M(\nu,T)}{A(\nu,T)}=f(\nu,T)$$

(3) 黑体辐射的单色辐出度的实验——近代物理学诞生的序幕。

问题：当时的理论无法解释实验曲线(图 4-4-7)的特点，无法找到实验曲线的函数关系式。维恩公式只是在短波方向与实验结果相符；瑞利-金斯公式只是在长波方向与实验结果相符，在短波方向单色辐出度(辐出功率)趋于无穷大，被称为“紫外灾难”(图 4-4-8)。

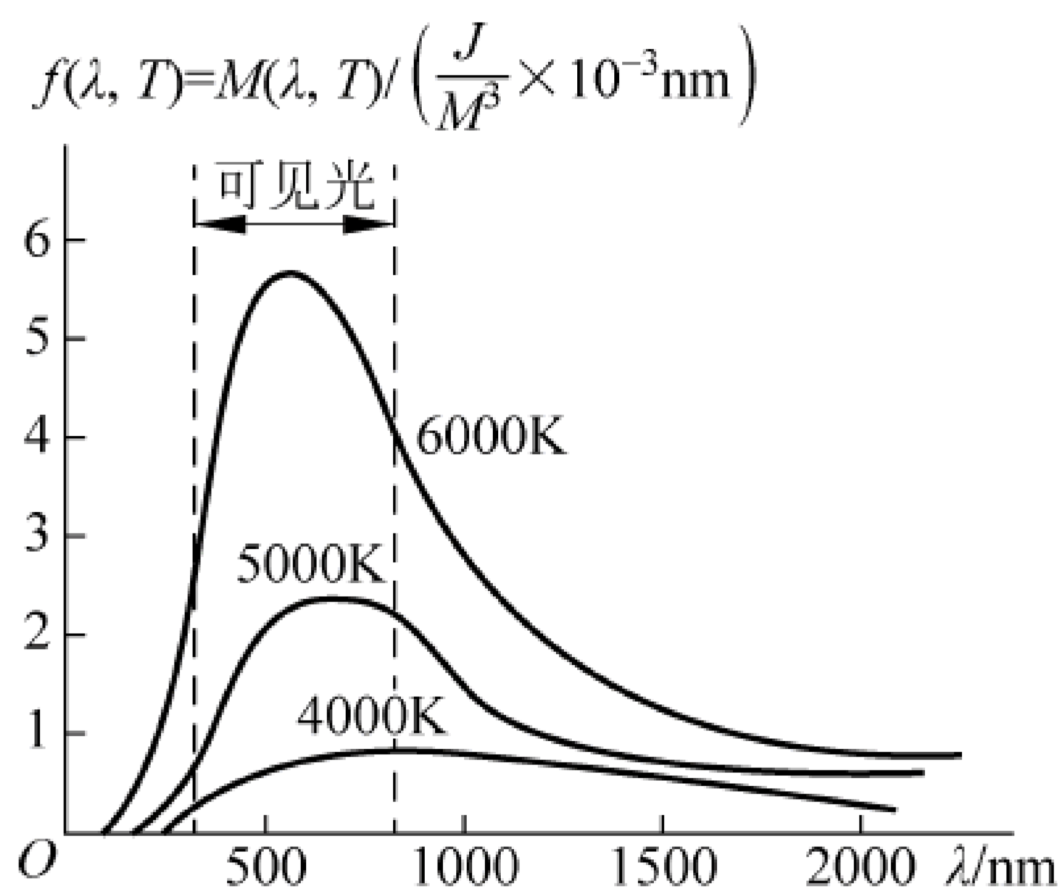



图 4-4-7 黑体辐射的实验曲线

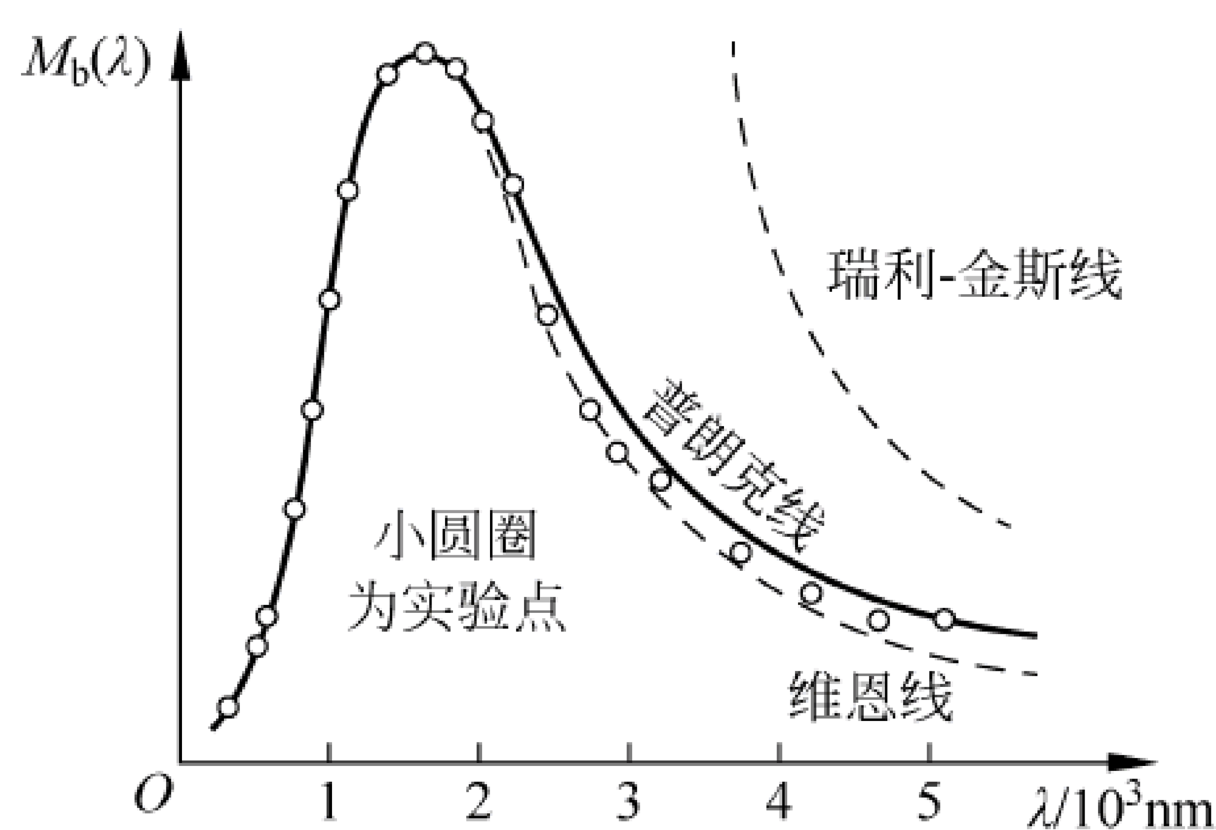



图 4-4-8 黑体辐射的实验曲线和理论曲线的比较

2）斯特藩-玻耳兹曼定律

文字表述：黑体的总辐出度 $M(\nu,T)$与绝对温度 $T$ 的四次方成正比。

数学表达：$M_0(T)=\int_0^{\infty}M(\nu,T)\mathrm{d}\nu=\int_0^{\nu}M(\lambda,T)\mathrm{d}\lambda=\sigma T^4$

3）普朗克辐射公式

(1) 普朗克假定

辐射物体中带电的线性谐振子(黑体空腔的谐振模式)与经典振子不同，它的能量不能

连续变化，只能处于某些特定的分立状态。

(2) 普朗克公式

$$f(\nu,T)=M_b(\nu,T)=\frac{2\pi}{c^2}\nu^2\overline{E_n}=\frac{2\pi h\nu^3}{c^2}\frac{1}{e^{h\nu/kT}-1} \tag{4-4-16}$$

(3) 简单讨论：公式与实验结果完全相符

短波辐射 $h\nu\gg kT$——维恩公式：

$$M_b(\nu,T)\approx\frac{2\pi h\nu^2}{c^2}e^{-h\nu/kT} \tag{4-4-17}$$

长波辐射 $h\nu\ll kT$——瑞利-金斯公式：

$$M_b(\nu,T)\approx\frac{2\pi\nu^2}{c^2}kT \tag{4-4-18}$$

普朗克公式圆满地解释了黑体辐射实验曲线的特点，普朗克分立能级的思想揭开了量子理论的序幕，从此物理学进入了第四次升华和大综合的阶段。

关于光学的详细讨论请参阅本书第8章、第9章和第10章。

## 4.5 原子物理学的理论体系

原子物理学，作为物理学的一个分支，主要研究物质结构的一个层次；这个层次介于分子和原子核两层之间，称之为原子，其研究内容主要包括原子的核外电子的能级特性，也包括分子物理和粒子物理的部分内容。

### 4.5.1 原子物理的研究对象和研究方法

研究对象：原子物理学属近代物理学范围，它是研究物质在原子层次内由什么组成，它们如何作用，发生什么样的运动形态的理论。即主要研究原子结构与性质及有关问题。它是关于物质微观结构的一门科学。

研究方法：从光谱及实验资料入手，提出假设，建立模型，然后再进行实验验证，最后形成理论。

### 4.5.2 原子物理学发展简史

**1. 原子结构模型的建立**

1897年J.J.汤姆孙发现电子，论证电子普遍存在，并确认它是各种原子的共同组成部分之后，对于在中性的原子内，正电荷和电子质量以及电子是如何分布的，成为摆在物理学家面前的首要问题。1904年汤姆孙提出原子的正电荷和质量均匀分布于原子体内、电子镶嵌在体内的“葡萄干圆面包模型”。1911年卢瑟福分析α粒子散射实验结果与汤姆孙原子模型的明显不一致，提出原子的有核模型，原子的正电荷和质量分布在中心很小的核内。原子的有核模型得到α粒子散射更为深入的实验研究支持而被普遍接受。但是在原子的有核模型中，电子绕核运动有加速度，根据经典电动力学，将不断向外辐射能量，电子将最终坍缩于原子核，因而原子是不稳定的；而且电子绕核运动发出连续谱也与实际上原子的线状光谱不符。这些事实表明，研究宏观现象确立的经典电动力学不适用于原子中的微观过程，因此需要进一步探索原子内部运动规律，建立适合于微观过程的原子理论。

**2. 原子物理学和量子力学**

1913 年玻尔在卢瑟福的原子有核模型基础上，结合原子光谱的经验规律，应用普朗克和爱因斯坦的量子概念，提出原子结构的新假设，建立玻尔氢原子理论，成功地解决了原子的稳定性问题，并说明了原子线状光谱的规律性。玻尔理论是原子理论发展的重要里程碑。1924 年德布罗意提出微观粒子具有波粒二象性，不久被实验证实，1926 年薛定谔、海森伯、玻恩、狄拉克等人建立微观粒子运动规律的量子力学。量子力学的建立为解决原子问题提供了锐利的武器，量子力学在阐明原子现象的种种问题中也逐步发展和完善，从而开创了近代物理的新时代。20 世纪 30 年代可称为原子物理的时代。原子物理学取得丰硕的成果，原子能级的结构和能级的精细结构、原子在外场中的能级结构、原子光谱规律、原子的电子壳层结构以及原子的深层能级结构和 X 射线标识谱等问题相继圆满解决，所获得的关于原子结构的种种知识成为了解分子的结构、固体的性质以及说明许多宏观现象和规律的基础。

原子物理学的新阶段在 20 世纪 50 年代末期，由于空间技术、空间物理和核试验的发展，不仅要求精确测定原子光谱的波长、研究原子的能级，而且对于谱线强度、跃迁概率、碰撞截面等也要求提供准确的数据，因此，要求对原子物理进行新的实验和理论探索。原子物理学的发展曾对激光的产生和激光技术的发展做出重大贡献。激光问世之后，应用激光技术研究原子物理学问题，实验精度有了很大提高，从而发现很多新现象和新问题。微波波谱学新的实验方法也成为研究原子能级结构的有力工具。因此，原子物理学的研究又重新成为很活跃的领域。原子碰撞研究已成为原子物理学的一个主要发展方向，研究课题非常广泛，涉及光子、电子、离子、中性原子等与原子和分子碰撞的物理过程，应用和发展了电子束、离子束、粒子加速器、同步辐射加速器、激光光源和各种能谱仪等测谱设备，以及电子、离子探测器、光电探测器和微弱信号检测方法。电子计算机的应用，加速了理论计算和实验数据的处理。原子光谱与激光技术的结合，达到了前所未有的高分辨率，利用激光高功率密度发展了非线性光学，饱和吸收、双光子吸收和多光子吸收等成为原子物理学中另一个十分活跃的研究方向。极端物理条件(高温、低温、高压、强场)下和特殊条件(高激发态、高离化态)下原子的结构和物性的研究也已成为原子物理研究中的重要课题。60 年代开始发展起来的将低能离子长时间约束在一个很小的空间范围内运动的离子存储技术，使人们可以从实验上近似得到孤立的、静止不动的单个带电粒子。近年来利用激光技术将中性原子降温减速并约束于空间很小范围内的原子囚禁技术取得重要的成果。这种存储技术正被应用于多种原子物理测量工作，测量精度更进一步提高，已成为量子电动力学理论最精确的检验手段之一，并渴望建立新的精度更高的光频标准。

原子物理学是其他基础科学和技术科学如化学、生物学、空间物理、天体物理、物理力学等的基础，激光技术、核技术和空间技术的研究也都要求原子物理学提供重要数据，因此研究和发展原子物理学至今仍有十分重要的理论和实际意义。

### 4.5.3 原子物理学的基本理论

**1. 理论框架**

众所周知，20 世纪以来物理学取得了惊人的成就：建立了相对论和量子理论，其适用于大到天体、小到夸克的许多领域，在其他学科的发展、生产、航天、信息技术应用等方面取得了前所未有的成就，而且处于不断的发展之中。所以，进入 20 世纪后期，原子物理学的内容逐渐增加量子力学的比重，原子物理学不再是原来主要进行对实验现象的解释，而是将波

粒二象性作为主线贯穿其中，作为大学物理学的重要组成部分，国内外有的教材体系干脆把这一部分内容合在一起，称为量子物理学，而不再沿用“原子物理学”的名称。

传统的原子物理学体系的结构
- (1) 原子的基本状况——质量、大小与结构
- (2) 原子的能级与辐射
- (3) 量子力学的基本原理
- (4) 碱金属与电子自旋
- (5) 多电子原子与辐射跃迁
- (6) 磁场中的原子
- (7) 原子的壳层结构与 X 射线
- (8) 分子结构与分子光谱
- (9) 原子核
- (10) 基本粒子

目前理论体系的结构——大学物理：量子物理学
- (1) 波粒二象性
- (2) 薛定谔方程
- (3) 原子中的电子——自旋、X 射线、激光与分子光谱
- (4) 固体中的电子——碱金属、量子统计与半导体
- (5) 核物理——核模型、放射性与核反应

**2. 原子的基本状况**

(1) 电子的发现

1897 年汤姆孙从如图 4-5-1 所示的放电管中的阴极射线发现了带负电的电子，并测得了 $e/m$ 值。1910 年密立根用油滴做实验发现了电子的电量值为 $e=1.602\times10^{-19}$C，从而求得电子质量是 $m_e=9.109\times10^{-31}\text{kg}=0.511\text{MeV}/c^2=5.487\times10^{-4}\text{u}$。

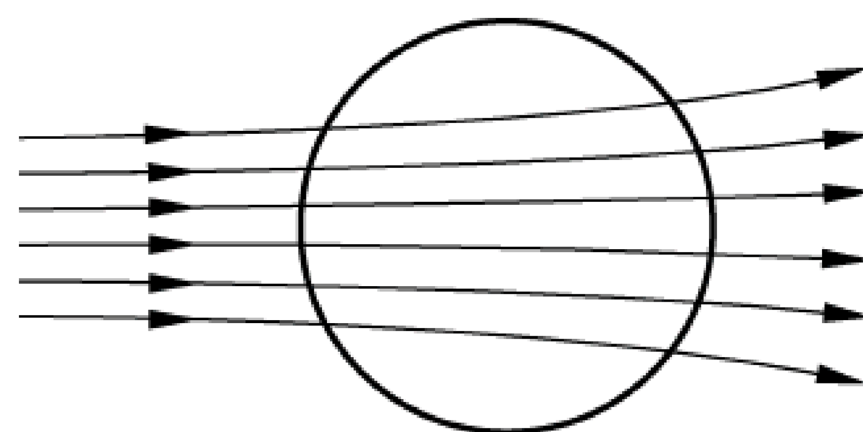


图 4-5-1　汤姆孙散射模拟

(2) α 粒子散射实验

目的：检验汤姆孙模型的正确性，汤姆孙模型如图 4-5-1 所示。

原理：带电粒子射向原子，探测出射粒子的角分布。α 粒子散射实验装置如图 4-5-2 所示。

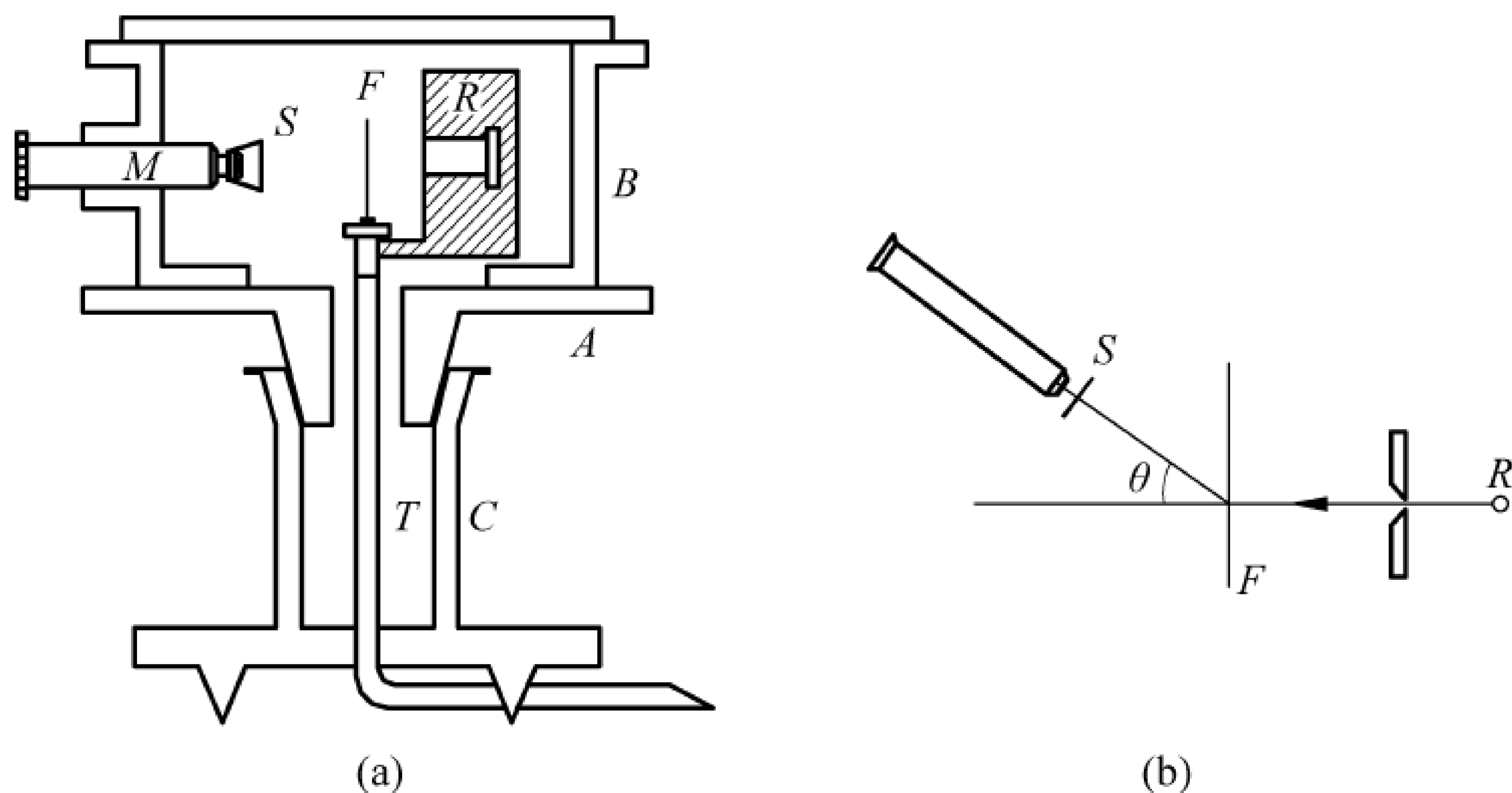



图 4-5-2　α 粒子散射实验装置和模拟实验

(a) 侧视图；(b) 俯视图

结果，大多数散射角很小，约 1/8000 散射大于 90°，极个别的散射角等于 180°。

(3) 原子核式结构模型——卢瑟福模型

原子序数为 $Z$ 的原子的中心，有一个带正电荷的核（原子核），它所带的正电量为 $Ze$，它的体积极小但质量很大，几乎等于整个原子的质量，正常情况下核外有 $Z$ 个电子围绕它运动。

(4) 核库仑散射角公式

动能为 $E_k$ 的 α 粒子从无穷远以瞄准距离 $b$ 射向原子核，在核库仑力作用下，偏离入射方向飞向无穷远，出射与入射方向的夹角 $\theta$ 称为散射角。其库仑散射公式：

$$b=\frac{a}{2}\cot\frac{\theta}{2} \tag{4-5-1}$$

式(4-5-1)中 $a=\frac{z_1z_2e^2}{4\pi\varepsilon_0E_k}$，$E_k=\frac{1}{2}mv^2$。这个过程称库仑散射，其中有两点假设：

其一，将卢瑟福散射看作是 α 粒子和原子核两个点电荷在库仑力作用下的两体碰撞，忽略原子中的电子的影响。

其二，在原子核质量 $M\gg m_\alpha$（α 粒子质量）时，可视为核不动，于是问题化为单质点 $m$ 在有心反平方库仑斥力作用下的运动问题。

(5) α 粒子散射实验的意义

① 实验解决了原子中正、负电荷的排布问题，建立了一个与实验相符的原子结构模型，使人们认识到原子中的正电荷集中在核上，提出了以核为中心的概念，从而将原子分为核外与核内两部分，并且认识到高密度的原子核的存在，在原子物理学中起了重要作用。

② α 粒子散射实验为人类开辟了一条研究微观粒子结构的新途径，以散射为手段来获得微观粒子内部信息的方法，为近代物理实验奠定了基础，对近代物理有着巨大的影响。

③ α 粒子散射实验还为材料分析提供了一种手段。

(6) 卢瑟福模型的困难

① 原子稳定性问题；

② 原子的同一性问题；

③ 原子的再生性问题；

④ 原子线状光谱问题。

**3. 玻尔理论及其局限性**

(1) 玻尔理论实验基础

① 黑体辐射——普朗克能量子；

② 光电效应——爱因斯坦光量子；

③ 氢原子光谱。

(2) 玻尔模型

① 基本假设；

② 圆轨道理论；

③ 氢原子光谱成因；

④ 非量子化轨道跃迁。

(3) 玻尔理论的困难

① 只能计算氢原子和类氢离子的光谱线频率，对于多于一个电子的氦原子，理论完全

不适用，且不能计算谱线的强度；

② 角动量量子化条件 $p_\varphi = n_\varphi \dfrac{h}{2\pi}$ 与现代实验结果不符只是人们的假设，无理论根据；

③ 轨道的概念不正确。

(4) 玻尔理论困难的原因

① 理论内在的不统一，不是自洽的。一方面提出了与经典理论完全矛盾的假设，另一方面又认为经典理论(牛顿定律、库仑定律)适用。所以不是一贯的量子理论，也不是一贯的经典理论，而是量子论 ＋ 经典理论的混合物。

② 没有抓住微观粒子的根本特性：波粒二象性。仍然把微观粒子看作经典理论中的质 点。更完整、更准确、应用面更广的关于原子的理论是1925年建立起来的量子力学。

**4. 量子力学**

1) 波粒二象性与德布罗意物质波

(1) 德布罗意假设

光电效应和康普顿散射告诉人们，光既有波的叠加——导致干涉、衍射的波动性，又具有质量、动量、能量和角动量——粒子的内禀属性的粒子性。光具有波粒二象性，并由爱因斯坦关系式描述为

$$\begin{cases} E = h\nu \\ p = h/\lambda \end{cases} \tag{4-5-2}$$

光的波粒二象性是否具有更深刻的普遍意义？德布罗意反向思考了这一问题，他指出："在整个世纪以来，人们在辐射理论上，比起关注波动的研究方法来，是过于忽视了粒子的研究方法；在实物理论上，是否发生了相反的错误呢？是否我们关于粒子的图像想得太多，而过分忽略了波的图像呢?"于是他将光的波粒二象性大胆地赋予了电子这样的实物粒子，即承认实物粒子也具有波粒二象性。

(2) 德布罗意关系式

德布罗意指出任何物体都伴随以波，不可能将物体的运动和波的传播分拆开来。这种波称为德布罗意物质波。德布罗意还给出了动量为 $p$ 的粒子所伴随波的波长 $\lambda$ 与 $p$ 的关系式为

$$\lambda = \frac{h}{p} \tag{4-5-3}$$

这就是著名的德布罗意关系式。

另外，自由粒子的能量 $E$ 和所伴随的波的频率 $\nu$ 之间的关系为 $E = h\nu$。

2) 薛定谔方程

(1) 自由粒子的薛定谔方程

薛定谔方程是量子力学的基本动力学方程，它在量子力学中的地位和作用相当于牛顿力学中的牛顿方程、电磁学中的麦克斯韦方程。薛定谔方程描述了量子系统状态的演化规律。

下面用一种直观的方法引出薛定谔方程。

考察质量为 $m$，动量为 $p$，能量为 $E = p^2/2m$ 的自由粒子的一维运动，它对应的德布罗意波是波矢为 $k$，圆频率为 $\omega$ 的平面波，即 $\psi(x,t) = \psi_o \mathrm{e}^{\mathrm{i}(kx-\omega t)}$，式中的 $k = 2n/\lambda$，$\omega = 2n\nu$。按照德布罗意关系式 $\lambda = h/p$ 和关系式 $E = h\nu$，一维自由运动的粒子的动量 $p_x$ 和能量 $E$ 与平面波波矢 $k$ 和圆频率 $\omega$ 有如下关系：$p_x = \dfrac{h}{\lambda} = \hbar k$ 和 $E = h\nu = \hbar\omega$。于是德布罗意平面

波可改写为

$$\psi(x,t)=\psi_o \mathrm{e}^{\mathrm{i}(p_x x-Et)/\hbar} \tag{4-5-4}$$

这个德布罗意波函数就是描述具有确定能量和动量的自由粒子运动的态函数。不难看出,若要从这个态函数中提取粒子的动能和动量信息,则必须用时间和空间坐标的微分算符作用其上方可给出,即

$$\begin{cases} \mathrm{i}\,\hbar\dfrac{\partial}{\partial t}\psi=E\psi \\ -\mathrm{i}\,\hbar\dfrac{\partial}{\partial x}\psi=p_x\psi \end{cases} \tag{4-5-5}$$

对于非相对论自由粒子能量-动量关系式

$$E=\frac{p_x^2}{2m} \quad 或 \quad \hbar\omega=\frac{(\hbar k)^2}{2m} \tag{4-5-6}$$

(2) 含时薛定谔方程

也可以通过如下算符作用在波函数 $\psi(x,t)$ 上得到

$$\mathrm{i}\,\hbar\frac{\partial}{\partial t}\psi(x,t)=-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t) \tag{4-5-7}$$

该式就是自由粒子一维运动的波方程,将其推广到三维情况,$E=p^2/2m$ 波动方程是

$$\mathrm{i}\,\hbar\frac{\partial}{\partial t}\psi(r,t)=-\frac{\hbar^2}{2m}\nabla^2\psi(r,t) \tag{4-5-8}$$

式中$\nabla^2=\left(\dfrac{\partial^2}{\partial x^2}+\dfrac{\partial^2}{\partial y^2}+\dfrac{\partial^2}{\partial z^2}\right)$是拉普拉斯算符。如果粒子在势场 $V(r,t)$ 中作三维运动,粒子的总能量是 $E=\dfrac{p^2}{2m}+V(r,t)$;相应的波方程应该是

$$\mathrm{i}\,\hbar\frac{\partial}{\partial t}\psi(r,t)=\left[-\frac{\hbar^2}{2m}\nabla^2+V(r,t)\right]\psi(r,t)\equiv\hat{H}\psi(r,t) \tag{4-5-9}$$

式中 $\hat{H}$ 称哈密顿算符,该式就是薛定谔方程,该方程是线性齐次方程,因而它保证了波函数(即态函数)的叠加性。

(3) 定态及定态薛定谔方程

如果势场不显含时间 $t$,即 $V=V(r)$,那么薛定谔方程成为

$$\mathrm{i}\,\hbar\frac{\partial}{\partial t}\psi(r,t)=\left[-\frac{\hbar^2}{2m}\nabla^2+V(r)\right]\psi(r,t) \tag{4-5-10}$$

仔细观察上式两边,不难发现方程的左边只含对时间微商的运算,右边只涉及对空间微商的运算,故可取分离变量式,即 $\psi(r,t)=\psi(r)f(t)$。并将其代入上式后,得到如下等式:

$$\frac{\mathrm{i}\,\hbar}{f(t)}\frac{\mathrm{d}f(t)}{\mathrm{d}t}=\frac{1}{\psi(r)}\left[-\frac{\hbar^2}{2m}+V(r)\right]\psi(r)=E \tag{4-5-11}$$

式中 $E$ 是既不依赖时间又不依赖空间坐标的常量(能量)。由式(4-5-11)分离出:$\mathrm{i}\,\hbar\dfrac{\partial}{\partial t}f(t)=Ef(t)$,它的解是 $f(t)=f_0\mathrm{e}^{-\mathrm{i}Et/\hbar}$。因此波函数具有形式(定态波函数)$\psi(r,t)=\psi(r)\mathrm{e}^{-\mathrm{i}Et/\hbar}$,其中波函数的空间部分满足

$$\left[-\frac{\hbar^2}{2m}+V(r)\right]\psi(r)=E\psi(r) \quad 或 \quad \hat{H}\psi=E\psi \tag{4-5-12}$$

式中 $\hat{H}=-\frac{\hbar^2}{2m}\nabla^2+V(r)$是哈密顿算符。

不是对任意的 $E$ 值才有解，而是只对一系列特定、分立值才有解，故这些特定的 $E$ 值可以用整数 $n$ 编序成 $E_n$，表明能量是量子化的。可见能量量子化自然蕴含在薛定谔方程中。方程

$$\hat{H}\psi_n=E_n\psi_n \tag{4-5-13}$$

正是能量本征方程。$E_n$ 是系统的一切可能的能量本征值，即常称的能级。$\psi_n$ 是本征值 $E_n$ 对应的本征函数或本征态。力学量能量用哈密顿算符表示；哈密顿算符有本征方程，通过求解该方程给出力学系统的一切可能的能量本征值及对应的本征函数，这是量子力学的基本假设。求解能量本征方程是量子力学最主要的任务。

**5. 原子中的电子**

(1) 电子自旋与轨道的相互作用能

自旋-轨道作用是原子内部磁相互作用的简称。由于电子有自旋磁矩 $\mu_s$，在电子为静止的坐标系上，原子实 $+NZ*e$ 绕电子旋转，并产生磁场 $B$。

自旋磁矩与该磁场作用，其作用能为

$$\Delta E_{sl}=-\boldsymbol{\mu}_s\cdot\boldsymbol{B} \tag{4-5-14}$$

(2) X 射线

性质：穿透性很强；使荧光屏发出荧光；使照相底片曝光；使气体电离；对动植物组织有刺激作用。

本质：X 射线具有光所具有的一切性质：反射、折射和偏振等，所以，X 射线从本质上来说是一种电磁波但其波长比通常的光波要短得多。

X 射线连续谱产生机制：快速电子射到阳极上，受到阳极中原子核的库仑场作用就会骤然减速；由此伴随产生的辐射，称为轫致辐射。

X 射线的特征(标识)谱：X 射线由 X 射线管产生。在 X 射线管中，从阴极发出的高速电子打到阳极上，由于电子能量很高，它能深入到阳极原子的内层，将内壳层电子之一击出原子之外，使原子电离，并在内壳层出现一个空穴。当邻近内壳层的电子跃迁到这个空穴时，就发射出波长很短的 X 射线。由于内壳层能级分立，所以，产生的 X 射线是线状谱。原子序数较大的元素，内壳层能级间隔就越大，发出的 X 射线的光子能量高，波长就短，所以，波长依次变化，不具有周期性。

每种元素都有一特定波长的线状光谱，即特征 X 射线谱，成为这种元素的标志。

(3) 康普顿效应

1923 年，康普顿在研究 X 射线经物质的散射实验中发现，散射的 X 线除有原入射波长成分外，还有波长较长的部分，其波长差随散射角 $\theta$ 而变化，如图 4-5-3 所示。

康普顿散射公式：经典瑞利散射不能解释散射波长随 $\theta$ 的变化，康普顿应用了爱因斯坦的光子概念，认为 X 射线经物质的散射是光子与外层电子(可视为自由电子)的碰撞过程。在碰撞中遵从能量和动量守恒，如图 4-5-4 所示。

动量守恒

$$\begin{cases}\dfrac{h\nu}{c}=mv\cos\alpha+\dfrac{h\nu'}{c}\cos\theta\\[2ex]0=-mv\sin\alpha+\dfrac{h\nu'}{c}\sin\theta\end{cases} \tag{4-5-15}$$

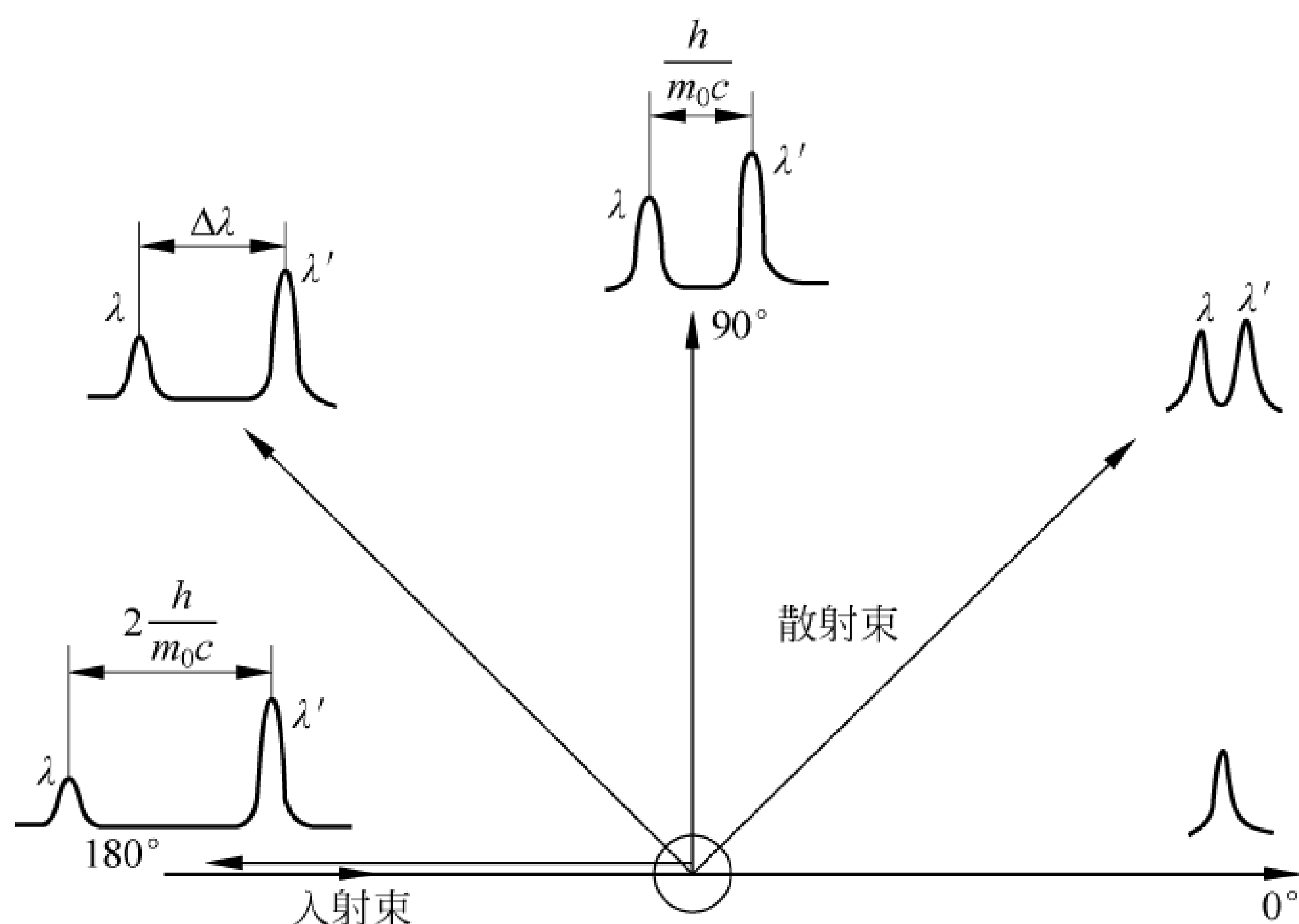



图 4-5-3　康普顿波长随角度的变化

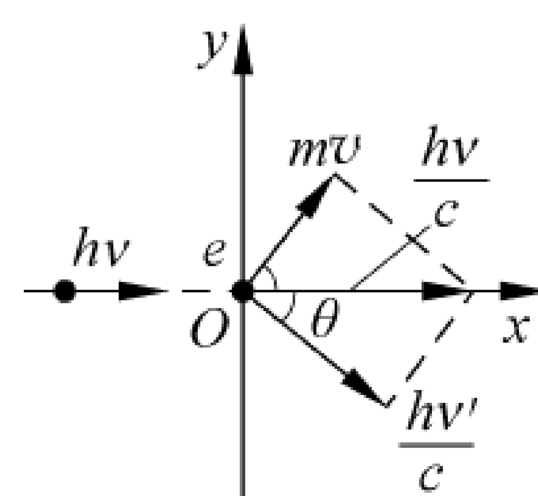



图 4-5-4　康普顿散射遵从能量和动量守恒

能量守恒

$$m_0c^2+h\nu=mc^2+h\nu' \quad (E_e=mc^2) \tag{4-5-16}$$

由以上三个方程可得康普顿波长变化量：

$$\Delta\lambda=\lambda_c(1-\cos\theta)=2\lambda_c\sin^2\frac{\theta}{2} \tag{4-5-17}$$

康普顿波长：$\lambda_c=\dfrac{h}{m_0c}=\dfrac{hc}{m_0c^2}=\dfrac{1.24}{511}\text{nm}=0.00243\text{nm}$

意义：用光子概念成功地解释了X射线的散射实验，这是继光电效应之后，再次证明光的粒子性。并指出像光子、电子这样的微观粒子也遵从动量守恒定律。

**6. 固体中的电子——碱金属**

(1) 碱金属原子光谱的实验规律

① 碱金属原子光谱具有原子光谱的一般规律性；

② 通常可观察到四个谱线系。

各种碱金属原子的光谱，具有类似的结构：主线系(也出现在吸收光谱中)，第二辅线系(又称锐线系)，第一辅线系(又称漫线系)，柏格曼系(又称基线系)。

(2) 碱金属原子的光谱项：$T(nl)=\dfrac{R}{n^{*2}}=\dfrac{R}{(n-\Delta l)^2}$

(3) 碱金属原子能级：$E_n=-hcT=-\dfrac{hcR}{n^{*2}}=-\dfrac{hcR}{(n-\Delta l)^2}$

(4) 碱金属原子能级特点

① 能量由$(n,\lambda)$两个量子数决定，主量子数相同，角量子数不同的能级不相同。

② $n$ 相同时能级的间隔随角量子数 $\lambda$ 的增大而减小，$\lambda$ 相同时，能级的间隔随主量子数 $n$ 的增大而减小。

③ $n$ 很大时，能级与氢的能级很接近，少数光谱线的波数几乎与氢的波数相同；当 $n$ 很小时，谱线与氢的谱线差别较大。

## 4.6　原子核物理学的理论体系

核物理学又称原子核物理学，是20世纪新建立的一个物理学分支。它与凝聚态物理、天体物理和原子物理等有着重要的联系。原子核物理的原理和技术在能源、新材料、环保、资源、考古、工农业、医学和生物学等诸多领域有着广泛的应用。核物理的研究和发展对社会的进步起着重要的推动作用。

原子核是比原子更深一个层次的物质结构。原子核物理学是研究原子核的性质，研究原子核的内部结构、内部运动、内部激发状态、衰变过程和裂变过程以及它们之间的反应过程的学科。

### 4.6.1　核物理学的发展历史

**1. 初期**

1896年，贝克勒尔(Antoine Henri Becquerel，1852—1908)发现天然放射性，这是人类第一次观察到的核变化。现在通常就把这一重大发现看成是核物理学的开端。此后的40多年，人们主要从事放射性衰变规律和射线性质的研究，并且利用放射性射线对原子核做了初步的探讨，这是核物理发展的初期阶段。

在这一时期，人们为了探测各种射线，鉴别其种类并测定其能量，初步创建了一系列探测方法和测量仪器。大多数的探测原理和方法在以后得到了发展和应用，有些基本设备，如计数器、电离室等，沿用至今。

探测、记录射线并测定其性质，一直是核物理研究和核技术应用的一个中心环节。放射性衰变研究证明了一种元素可以通过衰变而变成另一种元素，推翻了元素不可改变的观点，确立了衰变规律的统计性。统计性是微观世界物质运动的一个重要特点，同经典力学和电磁学规律有原则上的区别。

放射性元素能发射出能量很大的射线，这为探索原子和原子核提供了一种前所未有的武器。1911年，卢瑟福等人利用$\alpha$射线轰击各种原子，观测$\alpha$射线所发生的偏折，从而确立了原子的核结构，提出了原子结构的行星模型，这一成就为原子结构的研究奠定了基础。此后不久，人们便初步弄清了原子的壳层结构和电子的运动规律，建立和发展了描述微观世界物质运动规律的量子力学。

1919年，卢瑟福等又发现用$\alpha$粒子轰击氮核会放出质子，这是首次用人工实现的核蜕变(核反应)。此后，用射线轰击原子核来引起核反应的方法逐渐成为研究原子核的主要手段。

在初期的核反应研究中，最主要的成果是1932年中子的发现和1934年人工放射性核素的合成。原子核是由中子和质子组成的，中子的发现为核结构的研究提供了必要的前提。中子不带电荷，不受核电荷的排斥，容易进入原子核而引起核反应。因此，中子核反应成为研究原子核的重要手段。在30年代，人们还通过对宇宙线的研究发现了正电子和介子，这些发现是粒子物理学的先河。

20世纪20年代后期，人们已在探讨加速带电粒子的原理。到30年代初，静电、直线和回旋等类型的加速器已具雏形，人们在加速器上进行了初步的核反应实验。利用加速器可以获得束流更强、能量更高和种类更多的射线束，从而大大扩展了核反应的研究工作。此后，加速器逐渐成为研究原子核和应用技术的必要设备。

在核物理发展的最初阶段,人们就注意到它可能的应用,并且很快发现了放射性射线对某些疾病的治疗作用。这是它在当时就受到社会重视的重要原因,直到今天,核医学仍然是核技术应用的一个重要领域。

**2. 大发展时期**

20世纪40年代前后,核物理进入一个大发展的阶段。1939年,德国化学家哈恩(Hahn,Otto,1879—1968)等发现了核裂变现象;1942年,费米(Enrico Fermi,1901—1954)建立了第一个链式裂变反应堆,这是人类掌握核能源的开端。

在30年代,人们最多只能把质子加速到一百万电子伏特的数量级,而到70年代,人们已能把质子加速到四千亿电子伏特,并且可以根据工作需要产生各种能散度特别小、准直度特别高或者流强特别大的束流。

40年代以来,粒子探测技术也有了很大的发展。半导体探测器的应用大大提高了测定射线能量的分辨率。核电子学和计算技术的飞速发展从根本上改善了获取和处理实验数据的能力,同时也大大扩展了理论计算的范围。所有这一切,开拓了可观测的核现象的范围,提高了观测的精度和理论分析的能力,从而大大促进了核物理研究和核技术的应用。

通过大量的实验和理论研究,人们对原子核的基本结构和变化规律有了较深入的认识。基本弄清了核子(质子和中子的统称)之间的相互作用的各种性质,对稳定核素或寿命较长的放射性核素的基态和低激发态的性质已积累了较系统的实验数据。并通过理论分析,建立了各种适用的模型。

通过核反应,已经人工合成了17种原子序数大于92的超铀元素和上千种新的放射性核素。这种研究进一步表明,元素仅仅是在一定条件下相对稳定的物质结构单位,并不是永恒不变的。

天体物理的研究表明,核过程是天体演化中起关键作用的过程,核能就是天体能量的主要来源。人们还初步了解到在天体演化过程中各种原子核的形成和演变的过程。在自然界中,各种元素都有一个发展变化的过程,都处于永恒的变化之中。

通过高能和超高能射线束与原子核的相互作用,人们发现了上百种短寿命的粒子,即重子、介子、轻子和各种共振态粒子。庞大的粒子家族的发现,把人们对物质世界的研究推进到一个新的阶段,建立了一门新的学科——粒子物理学,有时也称为高能物理学。各种高能射线束也是研究原子核的新武器,它们能提供某些用其他方法不能获得的关于核结构的知识。

过去,通过对宏观物体的研究,人们知道物质之间有电磁相互作用和万有引力相互作用这两种长程的相互作用;通过对原子核的深入研究,才发现物质之间还有两种短程的相互作用,即强相互作用和弱相互作用。在弱作用下宇称不守恒现象的发现,是对传统的物理学时空观的一次重大突破。研究这四种相互作用的规律和它们之间可能的联系,探索可能存在的新的相互作用,已成为粒子物理学的一个重要课题。毫无疑问,核物理研究还将在这一方面做出新的重要的贡献。

核物理的发展,不断为核能装置的设计提供日益精确的数据,从而提高了核能利用的效率和经济指标,并为更大规模的核能利用准备了条件。人工制备的各种同位素的应用已遍及到理工农医各部门。新的核技术,如核磁共振、穆斯堡尔谱学、晶体的沟道效应和阻塞效应,以及扰动角关联技术等都迅速得到应用。核技术的广泛应用已成为现代化科学技术的

标志之一。

**3. 完善和提高时期**

20 世纪 70 年代,由于粒子物理逐渐成为一门独立的学科,核物理已不再是研究物质结构的最前沿。核能利用方面也不像过去那样迫切,核物理进入了一个纵深发展和广泛应用的新的更成熟的阶段。

在现阶段,粒子加速技术已有了新的进展。由于重离子加速技术的发展,人们已能有效地加速从氢到铀所有元素的离子,其能量可达到十亿电子伏每核子。这就大大扩充了人们变革原子核的手段,使重离子核物理的研究得到全面发展。

随着高能物理的发展,人们已能建造强束流的中高能加速器。这类加速器不仅能提供直接加速的离子流,还可以提供次级粒子束。这些高能粒子流从另一方面扩充了人们研究原子核的手段,使高能核物理成为富有生气的研究方向。

从核物理基础研究看,主要目标在两个方面:一是通过核现象研究粒子的性质和相互作用,特别是核子间的相互作用;再者是核多体系的运动形态的研究。很明显,核运动形态的研究将在相当长的时期内占据着核物理基础研究的主要部分。

## 4.6.2 核物理学的基本理论

**1. 原子核物理学的内容框架**

原子核物理学的内容主要分成三大块:原子核的结构与性质、原子核的转变、基本粒子,如图 4-6-1 所示。

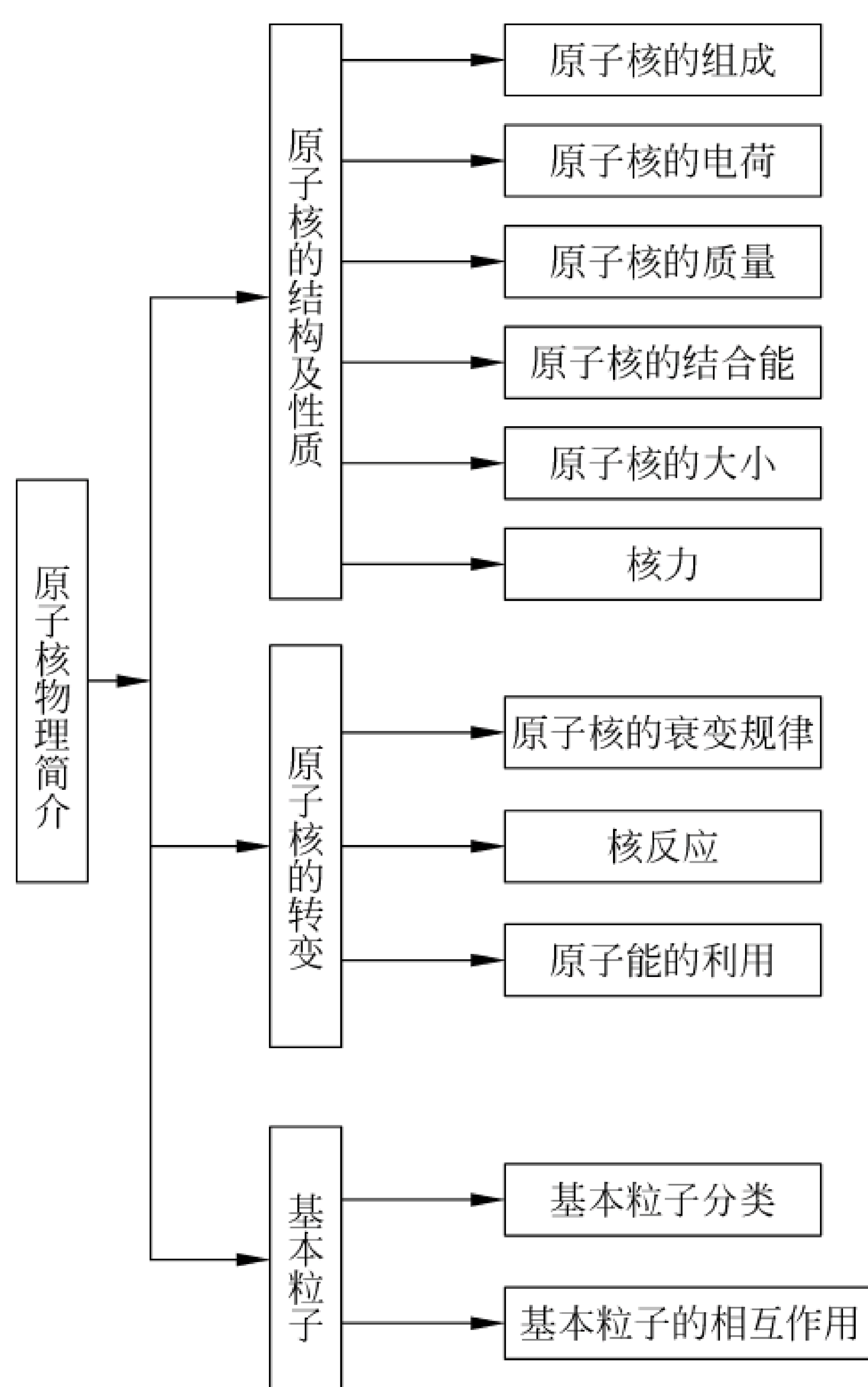



图 4-6-1　原子核物理学的内容框架

**2. 原子核的基本性质**

(1) 原子核的质子—中子模型

原子核由质子和中子组成。质子(p):+e,$m_p=1.007276\text{u}$;中子(n):电中性,$m_n=1.008665\text{u}$。质子和中子都是核子。在一切原子核中,核物质的密度是一个常数。

(2) 核自旋

原子核有自旋,自旋角动量为

$$P_I=\sqrt{I(I+1)}\,\frac{h}{2\pi} \tag{4-6-1}$$

其中,$I$ 为核自旋角动量量子数(核自旋)。核自旋可以是整数,也可以是半整数。

(3) 核磁矩

原子核有磁矩,以核磁子 $\mu_N=\dfrac{eh}{4\pi m_p}$为单位,即 $\mu_N=\dfrac{1}{1836.5}\mu_B=5.050\times10^{-27}\,\text{Am}^2$。

① 质子的磁矩:$\mu_p=2.792782\mu_N$;

② 中子的磁矩:$\mu_n=-1.913165\mu_N$。

中子的自旋角动量与磁矩方向相反。中子内部有一定的正负电荷分布,但正负电量相等,因此不显电性。

③ 原子核磁矩与核自旋角动量的关系

$$\mu_I = g_I \frac{e}{2m_p} P_I = g_I \frac{e}{2m_p} \sqrt{I(I+1)} \frac{h}{2\pi} = g_I \sqrt{I(I+1)} \mu_N \tag{4-6-2}$$

其中，$g_I$ 为原子核的 $g$ 因子。

(4)“核磁共振”测定核磁矩

$$\Delta E = -\mu_I B\cos\theta = -\mu_{I_z} B = -g_I M_I \mu_N B \tag{4-6-3}$$

原子核在磁场中有 $2I+1$ 个可能的能量，调节磁铁的励磁电流，使 $h\nu = \Delta E$，样品的原子核从磁场吸收能量，发生能级跃迁。核磁量子数的选择定则 $\Delta M_I = 0, \pm 1$，核能量改变 $\delta(\Delta E) = g_I \mu_N B$，当 $h\nu = g_I \mu_N B$，原子核从磁场中强烈吸收能量，称为核磁共振(见图 4-6-2)。

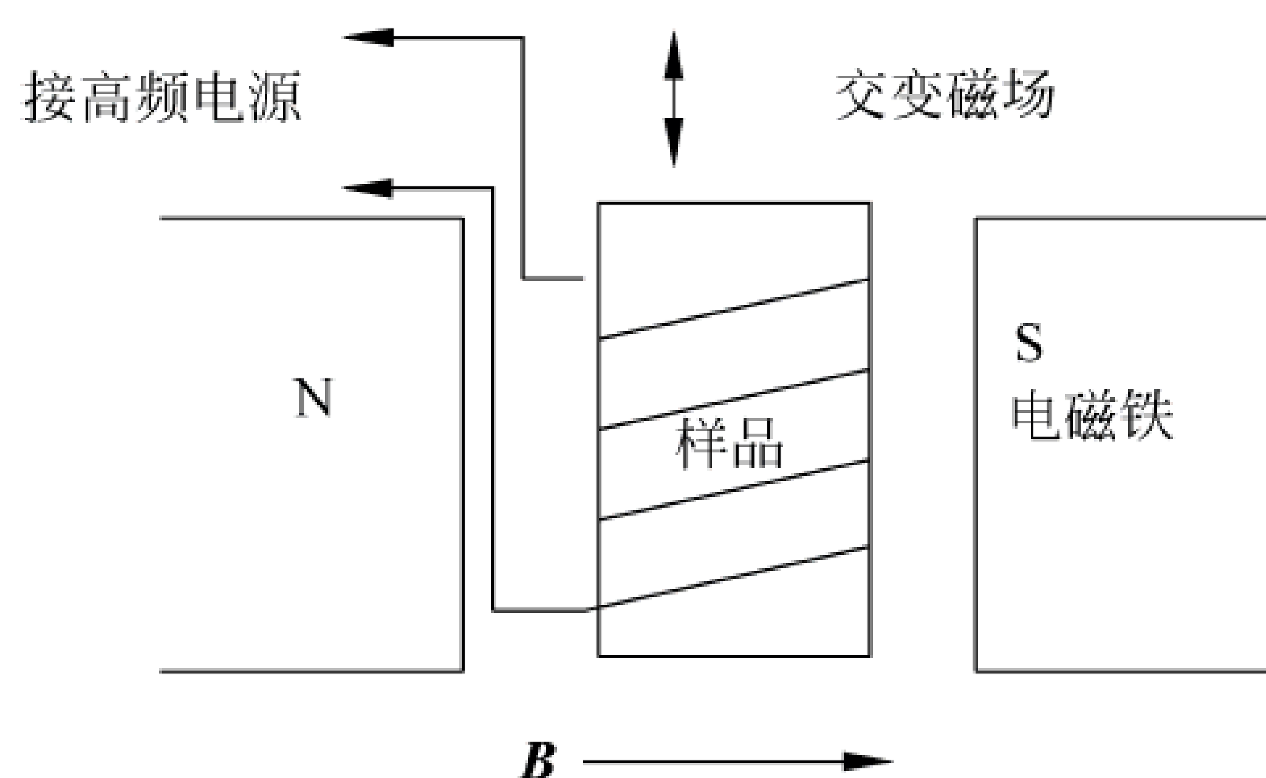



图 4-6-2 核磁共振的示意图

(5) 核力

核力是核子之间的相互作用力，抵消核子之间的较强的库仑排斥力。

核力的特征：

① 核力是比电磁力强得多的强相互作用力，主要是吸引力。

② 核力是短程力，核子间距离小于 $10^{-15}$ m 时才明显。

③ 核力与核子带电状况无关。

④ 核力具有饱和性。

(6) 原子核的结合能

① 原子核的结合能：由质子和中子形成原子核时所放出的能量。

② 质量亏损：原子核的质量总是小于组成该原子核的核子的质量之和，它们之间的差额称为原子核的质量亏损，计算公式为

$$\Delta m = [Zm_p + (A-Z)m_n] - M_x \tag{4-6-4}$$

其中，$M_x$ 为原子核的质量：$M_x = M - Zm_e$，$m_p = m_H - m_e$。所以得

$$\begin{aligned}\Delta m &= Z(m_H - m_e) + (A-Z)m_n - (M - Zm_e) \\ &= Zm_H + (A-Z)m_n - M\end{aligned} \tag{4-6-5}$$

核子在结合成原子核时，它们之间的核力作用使体系能量降低，从而释放出能量，相应质量减少了。

③ 平均结合能：原子核的结合能与原子核内所包含的总核子数的比值，$\bar{E}_0 = \dfrac{\Delta E}{A}$，平均结合能越大，原子核越稳定。

**3. 原子核的放射性衰变**

自然界中存在一些不稳定的原子核，这些原子核能自发地放射出一些射线，从一种状态转变为另一种状态，或从一种元素的原子核转变成另一种元素的原子核。这种现象称为原子核的放射性衰变。

(1) 放射性的发现

1896 年，法国物理学家贝克勒尔在研究物质的荧光现象时发现放射性。

1898 年，居里夫妇首先提炼出放射性同位素铀。

(2) 射线的性质

1899 年,卢瑟福等人用在垂直于射线方向加磁场的方法,对射线的性质进行了研究,发现了 α、β 和 γ 射线。

(3) 放射性发现的意义

放射衰变现象、电子、X 射线是 19 世纪末的三大重要发现,揭开了近代物理的序幕,提供了原子核内部运动变化的许多重要信息。

(4) 放射性衰变规律

放射衰变是一种自发的过程,其遵从统计规律:$N=N_0e^{-\lambda t}$,$N_0$ 是衰变前原子核,$N$ 是 $t$ 时刻现在的原子核数,$\lambda$ 是衰变常数。

(5) 三种放射性衰变

不稳定的原子核会自发地转变成另一种核而同时放出射线,这种变化叫放射性衰变。原子核放射性衰变发出三种射线:α 射线、β 射线、γ 射线。放射性衰变过程遵守电荷守恒、质量数守恒、能量守恒、动量守恒、角动量守恒。

① α 射线是 α 粒子流,是带正电的氦核,比如,${}^{226}_{88}Ra \longrightarrow {}^{222}_{86}Rn+{}^{4}_{2}He$。

② β 射线是高速运动的电子流,分 $\beta^+$ 和 $\beta^-$ 两种,比如,${}^{60}_{27}Co \longrightarrow {}^{60}_{28}Ni+{}^{0}_{-1}e+\tilde{\nu}_e$

β-衰变是原子核内中子转变成质子,同时放出一个电子和与电子相联系的反中微子:



$$ {}^{1}_{0}n \longrightarrow {}^{1}_{1}p+{}^{0}_{-1}e+\tilde{\nu}_e \tag{4-6-6} $$

β+衰变是原子核内质子转变成中子,同时放出一个正电子和一个中微子:



$$ {}^{1}_{1}p \longrightarrow {}^{1}_{0}n+{}^{0}_{1}e+\nu_e \tag{4-6-7} $$

③ γ 射线是光子流,是在 α 衰变或 β 衰变后形成新核时辐射出来的。

(6) 放射性衰变定律

① 衰变定律:$t$ 时刻样品中有 $N$ 个核,在 $dt$ 时间内有 $dN$ 个发生衰变,则 $-dN=\lambda N dt$,或

$$ \frac{dN}{N}=-\lambda dt \tag{4-6-8} $$

其中,$\lambda$ 表征衰变快慢的常数(衰变常数):$\lambda=\frac{-dN/dt}{N}$。当 $t=0$,$N=N_0$,则放射性衰变定律

$$ N=N_0e^{-\lambda t} \tag{4-6-9} $$

$t$ 时刻,每单位时间衰变的原子核数与该时刻原子核总数的比越大,衰变越快。

② 半衰期:原子核衰变到 $N=N_0/2$ 所需的时间。由 $\frac{1}{2}N_0=N_0e^{-\lambda T_{1/2}}$,得半衰期

$$ T_{1/2}=\frac{\ln 2}{\lambda}=\frac{0.693}{\lambda} \tag{4-6-10} $$

**4. 核结构模型**

核模型有壳模型、集体模型、滴模型等。

(1) 壳模型的主要实验依据——原子核存在幻数

壳模型的基本思想:原子核内每个核子处在其他核子所组成的平均势场中,其独立运动同时存在自旋-轨道相互作用,于是核子按一定顺序占据 $\nu lj$ 为定值的状态,形成按壳层分布的结构。

对于同一 $n_0$,可以有几组 $\nu$ 和 $l$ 的值。对最低能级的 $n_0=0$,此时只有一组:$\nu=1$,$l=$

0,即对 $n_0=0$ 的能级只有 1$s$ 一个状态。对 $n_0=1$ 的能级,只能 $\nu=1,l=1$ 和 $j=3/2,1/2$,因此,只能有 $1p_{1/2}$ 和 $1p_{3/2}$ 两个状态。对 $n_0=2$,可以有 $\nu=2,l=0$ 和 $\nu=1,l=2$ 和 $j=3/2,5/2$,这时包含 2$s$,$1d_{3/2}$ 和 $1d_{5/2}$ 三个状态,依次类推。

由于自旋-轨道耦合的引进,1$g$、1$h$ 和 1$i$ 能级劈裂得特别大,落入相邻的主壳层,得到了 50,82,126 等幻数,并构成一个个的主壳层。主壳层内的能级叫作支壳层。且质子和中子各自形成自己的壳层。

(2) 集体模型

① 集体模型的基本思想

其一,原子核中各核子既在各自轨道上独立运动,又在作集体运动。与壳模型相比,其势场有两个主要差别:由于原子核可以发生形变,势场不一定是球对称的;由于原子核有集体运动,势场不是静止而是变动的。

其二,单个核子在变动的势场中运动(也就是说,受集体运动的影响),但是单个核子的运动又会影响集体运动。

其三,集体运动要比单粒子运动缓慢得多。

② 集体运动的实验依据

其一,原子核具有电四极矩。

其二,核变形的存在是它存在集体运动的实验依据。

其三,偶偶核低激发能级的规律反映原子核具有集体运动。

③ 集体模型小结

其一,满壳层附近原子核平衡形状为球形,它反抗形变的能力很大,称为单粒子激发能级(双幻数附近)。

其二,离满壳层稍远(过渡区),核子在近似球形(略微变形)的平均场中进行独立运动。仍以球形为平衡形状,但抗形变能力减小。

其三,两个满壳层之间的原子核(大形变区)以旋转椭球为平衡形状。

④ 原子核激发方式有二

其一,单核子激发,能量较高;

其二,集体转动,或者集体振动加转动。

**5. 核反应**

放射性核衰变是不稳定核的自发转变,核反应是原子核在外界作用下的转变过程。研究核反应的重要目的之一是获取核能(裂变能、聚变能)。

(1) 核反应的形式:核裂变(重核分裂为轻核)和核聚变(轻核聚合为重核)。

(2) 守恒定律

核反应遵守质量数守恒、电荷数守恒、能量守恒、动量守恒、角动量守恒以及宇称守恒。

(3) 反应能 $Q$

即核反应过程中放出或吸收的能量等于反应前后体系的动能之差。

$M_x+M_X>M_y+M_Y$,则 $Q>0$,为放能反应;

$M_x+M_X<M_y+M_Y$,则 $Q<0$,为吸能反应;

$M_x+M_X=M_y+M_Y$,则 $Q=0$,为弹性散射。

(4) $Q$ 方程

$$Q=E_y\left(1+\frac{M_y}{M_Y}\right)-E_x\left(1-\frac{M_x}{M_Y}\right)-\frac{2\sqrt{M_xM_yE_xE_y}}{M_Y}\cos\theta$$

$$=E_y\left(1+\frac{A_y}{A_Y}\right)-E_x\left(1-\frac{A_x}{A_Y}\right)-\frac{2\sqrt{A_xA_yE_xE_y}}{A_Y}\cos\theta \tag{4-6-11}$$

(5) 核反应截面

假定入射粒子束强度为 $I$(单位时间通过单位面积的粒数)，打到一个靶上的横截面是 $S$，也就是说单位时间打到靶上的粒子为 $n=IS$，靶单位体积中有 $N$ 个原子核，靶的厚度为 $h$，$h$ 比较小，以前后各原子核不会互相遮蔽，这样入射粒子将看到全部 $NSh$ 个原子核。设想每个原子核提供一个截面 $\sigma$，凡是落入这个范围的入射粒子就和核发生反应，单位时间共发生了 $\Delta n$ 个反应，则显然

$$\sigma=\frac{\Delta n}{INSh}=\frac{\text{出射粒子数(反应粒子数)}}{\text{入射粒子数}\times\text{单位面积靶核数}} \tag{4-6-12}$$

反应截面实质上是单位强度的入射粒子束和一个靶核发生反应的概率。

### 4.6.3　基本粒子

**1. 基本粒子的分类**

(1) 规范粒子：是传递相互作用的媒介。种类有：光子 $\gamma$、胶子 g、中间玻色子 $W^{\pm}$ 和 $Z^0$。

(2) 轻子：参与弱相互作用，带电的轻子也参与电磁作用。

(3) 强子：参与强相互作用的粒子，也参与弱相互作用和电磁作用。

包括共振态粒子，到目前为止，人们已经发现了 200 多种基本粒子。

**2. 基本粒子间的相互作用**

可分为四种：引力(万有引力)、电磁作用、强作用和弱作用。

**3. 层子(或夸克)模型**

1964 年盖尔曼(Murry Gell-Mann，1929—2019)等人提出强子是由夸克组成的模型。1965 年我国理论工作者提出类似的层子模型。

基本思想：所有的重子都是由三个层子或夸克组成，反重子则由三个反层子或反夸克组成，介子是由一个层子和一个反层子构成。

关于基本粒子的一般讨论，请读者参阅本书 1.1 节和 1.2 节。

## 学习资料

### 一、中国近现代著名科学家(物理学)简介

竺可桢(1890—1974)，又名绍荣，字藕舫，浙江上虞人。汉族，当代著名的地理学家、气象学家和教育家，中国近代地理学的奠基人。竺可桢被公认为中国气象、地理学界的“一代宗师”。



竺可桢毕业于伊利诺伊大学农学院、哈佛大学。其代表作品：《竺可桢文集》和《竺可桢日记》。竺可桢是毕生为国“求是”的气象事业和物候学的开拓者，1908 年入上海复旦公学求学，1910 年公费留美入伊利诺伊大学农学院学习。1913 年夏毕业后转入哈佛大学研究院地理系专攻气象，1918 年以题为《远东台风的新分类》(*A NEW CLASSIFICATION OF THE TYPHOONS OF THE FAR EAST*)的论文获得博士学位。1920 年秋应聘南京高等师范

学校。1927 年任国立东南大学地学系主任，1928 年任中央研究院气象研究所所长，建国前他先后执教于武昌高等师范学校、国立东南大学和国立中央大学。1933 年 4 月，竺可桢与翁文灏、张其昀共同发出成立中国地理学会的倡议，学会于翌年成立。1936 年 4 月，他担任浙江大学校长，历时 13 年，被尊为中国高校四大校长之一，1962 年在南京大学建校 60 年之际题词“东南学府，为国之先。男女同校，唯此首创。外御强敌，内抑豺狼。天下有道，黉舍乃昌”。1949 年为了婉拒蒋介石赴台湾的邀请，辞去了浙大的所有职务。中华人民共和国成立后，担任中国科学技术协会副主席，中国气象学会理事长、名誉理事长，中国地理学会理事长等职。他对中国气候的形成、特点、区划及变迁等，对地理学和自然科学史都有深刻的研究。他在气象学、气候学、地理学、物候学、自然科学史等方面的造诣很深，而物候学也是他呕心沥血做出了重要贡献的领域之一。我国现代物候学的每一个成就都是和他的工作分不开的。他始终从科学的视角，关注着中国的人口、资源、环境问题，是“可持续发展”的先觉先行者。2008 年在由中国科学技术协会组织的评选中，与袁隆平、茅以升等一起获评中国十大科技传播优秀人物。

## 二、诺贝尔物理学奖简介

**1921 年诺贝尔物理学奖**。发现了光电效应的规律。获奖者：A. 爱因斯坦(德国)。评价：这次的颁奖或许是评委会最为艰难的一次，因为尽管爱因斯坦的相对论早已名扬天下，但是评委会却始终难以理解这一理论。但是如果爱因斯坦没有获奖，那么将会是一个更大的玩笑。所以就用光电效应的文章作为颁奖理由。其实就这一篇文章而言也完全足以获得物理学的至高荣誉。在这篇文章中他明确提出了光子概念。比起普朗克的含糊其辞，爱因斯坦前进了一大步。当然要是和他在相对论的成就相比就不算什么了。至于相对论大家喜欢讨论 $E=mc^2$，不少科普书都对此公式推崇备至，包括《时间简史》。其实大家只要记住光速是永恒不变的就可以了，无论你是在多快的地方去测量它都不会发生变化。通过这一基本假设推导出了一个新的世界。$E=mc^2$ 只是其中的一个推论。

**1922 年诺贝尔物理学奖**。原子结构和原子辐射的研究。获奖者：N. 玻尔(丹麦)。评价：作为 20 世纪最伟大的科学家之一，他和爱因斯坦的连续获奖不知是否出于巧合。玻尔可以说对量子力学做出了开创性的贡献，他提出了一个量子化的原子模型，认为电子只能在某几个能级中运动，跃迁对氢原子谱线作了很好的解释；还有他所创建的哥本哈根学派对物理学产生了巨大影响，培养出了一批又一批的大师。当然他的理论不是没有问题，薛定谔曾经问他：既然电子只能在这几个轨道中运动，那么从两者间跳变时它在哪个地方？

**1923 年诺贝尔物理学奖**。基本电荷和光电效应的研究。获奖者：R. A. 密立根(美国)。评价：密立根的名字大家不会陌生，高中时密立根油滴实验可是经常的考点，不过用油滴实验来测量分子直径似乎有些“大材小用”。密立根通过此实验测量电荷那可是相当艰苦的工作。他做了几千次测量，一个油滴要盯住几个小时，可见其艰苦。这还不算，由于在此极端情况下数学上的一些公式已不适用，还要自己推导另外的公式。所以密立根可以说无愧为 20 世纪最伟大的实验物理学家。在他的论文中对自己实验有些误差也如实坦诚，这告诉我们一个实验物理学家就要忠于自己的数据。不过由于这些误差正好是电荷的 1/3，所以很有可能和夸克模型有关。

**1924 年诺贝尔物理学奖**。X 射线光谱学方面的研究和发现。获奖者：K. M. G. 西格巴恩(瑞典)。评价：自巴克拉发现元素的次级 X 射线标识谱后不少科学家投入到了这项研究

中。其中西格巴恩成就最高。由于他发明了分辨率极高的光谱仪，所以他对 X 射线的谱线研究最深入。他分析并确定了 92 种元素的原子所发射的标识 X 射线，对各种谱线也作了深入的分析。另外从伦琴开始，人们一直试图证明 X 射线是一种波长短的电磁辐射。1924 年，西格巴恩用棱镜演示 X 射线的折射获得成功，从而完成了这一历史使命。

**1925 年诺贝尔物理学奖**。发现原子和电子的碰撞规律。获奖者：G. 赫兹（德国），J. 弗兰克（德国）。评价：这个实验用电子与水银蒸汽原子碰撞，一般发生弹性碰撞（即无能量损失），但能量达到一定程度后电子会损失某些分立的能量（4.9eV 或整数倍），说明了汞（水银）原子的电子存在能级且只吸收一些分立的能量。证明了玻尔原子模型。

**1926 年诺贝尔物理学奖**。研究物质结构的不连续性，特别是发现沉积平衡。获奖者：J. B. 佩兰（法国）。评价：最著名的工作是他对布朗运动的研究证实了分子的客观存在。布朗运动，是悬浮在液体中微粒的一种无规则运动。现在人们知道，布朗运动是由周围分子对这些微粒不断的、无规则的碰撞引起的。通过对布朗运动规律的研究佩兰又得出了测定阿伏伽德罗常数的方法。佩兰的研究结束了关于分子是否真实存在的长期争论。此争论可以上溯到古希腊时期亚里士多德和德莫克里特关于物质是无限可分还是由基本粒子构成的争论。正如彭加勒所说："佩兰确定原子数目的杰出工作，宣告了原子学说的胜利，……化学家的原子，现在是一个客观实体了。"

**1927 年诺贝尔物理学奖**。发现康普顿效应，通过蒸气凝结观察带电粒子径迹的方法。获奖者：A. H. 康普顿（美国）、C. T. R. 威尔孙（英国）。评价：康普顿发现散射后的 X 射线包含两种不同的波长成分：一种和入射 X 射线波长相同，称为不变线；另一种则大于入射 X 射线的波长，称为变线。这就是康普顿效应。为了对实验观察到的变线波长给予合理的解释，康普顿于 1923 年提出了一种新的量子理论。他认为，在康普顿效应中，一光子与一个自由电子发生了保持能动量守恒的弹性碰撞（只有粒子才能碰撞）。首次用实验证明了光的粒子性。而威尔孙发明了能观测粒子轨迹的云雾室，并观察到康普顿效应后会有一个反冲电子。这是一项非常重要的发现：正电子、$\mu$ 子、$K_0$ 介子和 $\Xi$ 超子等都是通过拍摄它们在云雾室中的径迹而发现的。

**1928 年诺贝尔物理学奖**。从事热离子现象的研究，特别是发现理查森定律。获奖者：理查森（英国）。评价：20 世纪初无线电可以说是最重要的技术创新，马可尼和布劳恩早在 1909 年就获奖。而理查森对热离子现象的研究对无线电发射有重要的指导意义。它表明加热金属丝发射电子只是与金属丝本身有关，与周围物质无关。为此他整整花了 15 年时间去证实他的理论，比法拉第还多 5 年。他的获奖名至实归。

**1929 年诺贝尔物理学奖**。发现物质波。获奖者：德布罗意（法国）。评价：继爱因斯坦提出光子说认为光具有波粒二象性后，一个法国年轻人走得更远，他提出了一切物质都具有波粒二象性。并提出了波长的计算关系。德布罗意在他的博士论文中发表了他的观点并获得了诺贝尔奖。更值得一提的是他的理论距离薛定谔方程只有一步之遥。量子力学的第一个基本方程已经呼之欲出。

**1930 年诺贝尔物理学奖**。从事光散方面的研究，发现喇曼效应。获奖者：C. V. 喇曼（印度）。评价：在 X 射线的康普顿效应发现以后，人们就开始想可见光是否能观察到这些现象。而喇曼就观察到了光在色散后部分光频率改变的性质并为光的量子理论提供了新的证据。喇曼是第一位获得诺贝尔物理学奖的亚洲科学家。他因为生病没有作博士论文，所以，只能去财政部（当时规定在科学文化界任职的印度人必须有英国博士学位）。后来他在

实验室研究并取得一系列成果,后被一所大学破例聘为教授,并取得了辉煌成果。此外,喇曼还是一位教育家,他从事研究生的培养工作,并将其中很多优秀人才输送到印度的许多重要岗位。

## 思考与练习

1. 大学物理学的理论是如何划分的,划分的依据是什么?
2. 如何理解高等数学在物理学理论发展过程中的巨大作用?
3. 物理定律与定理有何不同?
4. 经典力学的基本定律有哪些?
5. 如何理解热力学与统计力学的联系与区别?
6. 试从辩证科学思维的角度论述电磁学的发展与建立?
7. 光学研究的理论基础是什么?
8. 原子物理学和原子核物理学有哪些理论和应用成就?

# 第5章

# 牛顿定律与力学守恒定律

牛顿运用逻辑推理、实验探究和数学演绎等方法，对前辈科学家留下的力学、天文学大量的零散知识和研究成果进行了系统的升华和提炼，并赋予它们以新的逻辑形式和严谨的数学形式。正是在这一系列的创新活动中，牛顿发明了解决精确计算问题的工具——微积分，体现了牛顿的非凡创造力，牛顿的力学和数学成就标志着近代精确科学的诞生。牛顿三大定律和万有引力定律是解决一切宏观力学问题的根本依据，是整个力学的基础。

第4章系统地讨论了力学体系的建立和发展过程，本章较为详细地讨论了处于力学核心地位的牛顿三定律的物理含义、相互关系与历史地位，以及力学中常见的一些应用问题。

## 5.1　力的概念及其分类

力学是研究系统在受力的作用时的行为的科学。按照系统的类型和所涉及的现象，力学划分为若干个分支。与其他学科的一个重要的区别是力学的划分以系统的空间尺寸为划分依据。尺寸足够大的宏观系统可以由经典力学的牛顿定律合适地描述。所谓的经典力学，就是研究通常尺寸的物体在受力下的形变，以及速度远低于光速的运动过程的一门自然科学。经典力学在很多情况下也简称为力学，它是物理学、天文学和许多工程力学的基础，机械、建筑、航天器和船舰等的合理设计都必须以宏观力学为基本依据。

本节探讨力的概念形成的历史过程以及力的分类。

### 5.1.1　力学发展的历史

古代人从对日、月运行的观察和弓箭、车轮等的使用中，了解一些简单的运动规律，如匀速的移动和转动。人们在日常生活和劳动中使用杠杆、打水器具，逐渐认识了物体受力及平衡的情况。力是物体间的一种相互作用，这是人们最早由朴素的感性认识得出的，实际上也反映出人们对力的本质特征的感悟。力的概念最初是人类在劳动中通过肌肉紧张的感觉而产生的，后来被科学家不断归纳、概括和抽象，成为物理学最基本的概念。

在公元前480—380年，中国的墨翟(约前478—392)及其子弟所著的《墨经》一书中，对运动和力有明确的定义。在力的概念及许多力学知识方面，墨家理论走在同时代的前列，闪耀着中华民族智慧的光辉。

墨家考察了人体对周围环境的作用，看到人们通过肌肉的动作，如举、持、掷、踢、蹬，可以使别的物体发生位置移动，从而总结出肌肉力的概念。《墨经》第二十一条指出："力，刑之所以奋也。"这里"刑"同"形"，指人体、物体；"奋"，指克服阻抗而运动。该条阐述了"力"的含义：即力是使人和物由静到动的根本原因。

墨家还把力和"重"联系起来，把"重"看成一种力，可以说这是人类对力的最早的理性认

识。《墨经》第一百二十七条的《经说》还指出："凡重，上弗挈，下弗收，旁弗劫，则直下。"意思是说，具有重量的物体，自由落体下落时，其轨迹必定是竖直向下的。这说明墨家对重量产生的力效果研究得十分深入。

墨家对阻力、惯性、浮力以及墙体和梁的受力等常见力学现象都有精细的分析。对力的平衡、杠杆、滑轮、轮轴和斜面的力学知识也有深入的理解，并指出"相衡，则本短标长"，说明在称量重物时，要想与砝码平衡，就要调整重臂"本"和力臂"标"，这一记录比阿基米德早二百多年。

在西方，人们对力学的研究更加深入具体，而且一直持续到现代，不断取得进步。我们都知道，亚里士多德认为物体的运动只有在外界的不断作用下才能产生，也就是说，他认为力是产生运动的原因。但这种说法解释不了抛射运动，于是出现了冲力说，认为上帝创世之初就赋予天体一种"冲力"。这是一种不随时间流逝的动力，这种动力可以维持物体永远运动下去。因此，运动的物体一般并不需要经常有个和它接触的推动者。

著名的意大利艺术家、物理学家和工程师达·芬奇是文艺复兴时期首先跳出中世纪脱离实际和实验的烦琐论证的人之一，他认为实验和数学运用对于解决力学问题有巨大意义。他应用力矩法解释了滑轮的工作原理；应用虚位移原理的概念来分析起重机结构中的滑轮和杠杆系统；在他的一份草稿中，他还分析了铅垂力的分解；研究了物体的斜面运动和滑动摩擦阻力，最先得出了滑动摩擦阻力同物体的摩擦接触面大小无关的结论。

伽利略在实验研究和理论分析的基础上，最早阐明自由落体运动的规律，提出加速度的概念。牛顿继承和发展前人的研究成果(特别是开普勒的行星运动三定律)，提出物体运动三定律。伽利略、牛顿奠定了动力学的基础。牛顿运动定律的建立标志着力学开始成为一门科学。

牛顿通过对运动的深入研究，总结出力具有大小、方向和作用点三要素，并发现了牛顿三大定律。他从根本上把握了惯性定律、力的大小和动量的变化率成正比以及作用力与反作用力等大反向等规律，从此，人们对力由定性的归纳抽象逐渐转入定量的分析研究。

牛顿认为，所谓力就是使物体运动状态发生变化的原因。他说明，作为定量的定义，力是和动量的变化率成正比的，即力是物体对物体的作用，力可以用受力物体的动量变化率来度量。此后弹性力学和流体力学基本方程的建立，使得力学成为拥有完整体系的独立学科。

到20世纪初，在实际应用中，流体力学和固体力学与数学理论互相结合，使力学蓬勃发展，而且许多新的力学分支理论的创立，解决了工程技术中大量的关键性问题。

尽管人们知道宏观力是物体间的相互作用，是物体运动状态发生变化的原因，但人们无法进一步弄清力究竟是一种什么样的作用，力产生具体原因是什么。所以，力的精确定义问题至今仍未彻底解决。至今，各种力的概念仍然难以被人理解、物理意义仍不确切、其本质也没有被揭示出来，请看下面关于自然界的四种基本力的归纳。

### 5.1.2 自然界的四种基本力

通过大量的研究，到目前为止，人们发现自然界中各种各样的相互作用都来源于四种基本力，即万有引力、电磁力、强力和弱力。人们对这四种基本力的物理意义、本质以及四种基本力的统一性问题仍然处于探索之中。

**1. 万有引力(gravitational force)**

牛顿万有引力定律的表述：存在于一切物体间的相互吸引力。

数学形式：

$$\boldsymbol{F}=-G_0\frac{m_1m_2}{r^2}\boldsymbol{e}_r \tag{5-1-1}$$

其中，$m_1$ 和 $m_2$ 为两个质点的质量，$r$ 为两个质点间的距离，$G_0=6.67\times10^{-11}\text{m}^3\text{kg}^{-1}\text{s}^{-2}$ 叫作万有引力常量。

万有引力是普遍存在的力，大到天体之间、小到基本粒子之间都有相互的吸引力。其力程如表 5-1-1 所示。

**2. 电磁力（electromagnetic force）**

存在于静止电荷之间的电性力以及存在于运动电荷之间的电性力和磁性力，它们本质上是相互联系的，总称为电磁力。

分子或原子都是由电荷系统组成，它们之间的作用力本质上是电磁力。例如：物体间的弹力、摩擦力，气体的压力、浮力、黏滞阻力。

**3. 强力（strong interaction）**

亚微观领域，存在于核子、介子和超子之间、把原子内的一些质子和中子紧紧束缚在一起的一种力。

**4. 弱力（weak interaction）**

亚微观领域内的另一种短程力是导致 $\beta$ 衰变放出电子和中微子的重要作用力。

具体来说，凡是有质量的物质都具有万有引力，凡是带有电荷的物质都具有电磁力，而在原子核内部的基本粒子（包括各种夸克）之间则存在着强相互作用力和弱相互作用力。

表 5-1-1 为四种基本力的比较。

**表 5-1-1 四种基本力的比较**

| 类型 | 源 | 相对强度 | 作用距离 |
|---|---|---|---|
| 万有引力 | 质量 | $10^{-38}$ | 长（无限） |
| 电磁力 | 电荷 | $10^{-2}$ | 长（无限） |
| 强力 | 强子 | 1 | 短（$10^{-15}$ m） |
| 弱力 | 所有粒子 | $10^{-15}$ | 短（$10^{-17}$ m） |

这四种力又可以分为两类：万有引力和电磁力在物体相距较远时仍发挥作用，叫做长程力。万有引力在天体层次的运动中起重要作用。电磁相互作用不仅在宏观现象并且在微观现象中发挥作用。另一类是强相互作用与弱相互作用，它们的作用距离很短，叫短程力。短程力只有在微观现象中才发挥明显作用。

许多物理学家，特别是优秀的天才的物理学家，都试图把上述四种自然基本力统一成一种基本力，如果把前三种力统一起来称为“大统一理论”（grand unified theory），若把四种力统一起来则称“万物理论”（theory of everything）。目前还有两大理论仍在研究之中：超对称理论（supersymmetry）以及超弦理论（superstring）。

## 5.1.3 力的分类

在日常生活中，由于力的普遍存在，人们对力的定性感觉都非常熟悉。但是，由于力的多样性，它没有一个统一的划分标准。按照力不同的表现形式、作用等，物理学对力按照以下几种常见的方法进行分类。

**1. 根据力的效果、性质分类**

在力学研究中，我们经常根据力的效果给力命名，如拉力、压力、支持力、动力、阻力等。

从力的性质上分则有重力、弹力、摩擦力、分子力、电磁力等,称为常见力。根据力的效果、性质,其分类如图 5-1-1 所示。效果不同的力,性质可以相同。例如压力和支持力都是弹力,只是效果不同。性质不同的力,效果可以相同。例如不论是什么性质的力,只要效果是加快物体运动的,就可以称它为动力;效果是阻碍物体运动的,就可以称它为阻力。

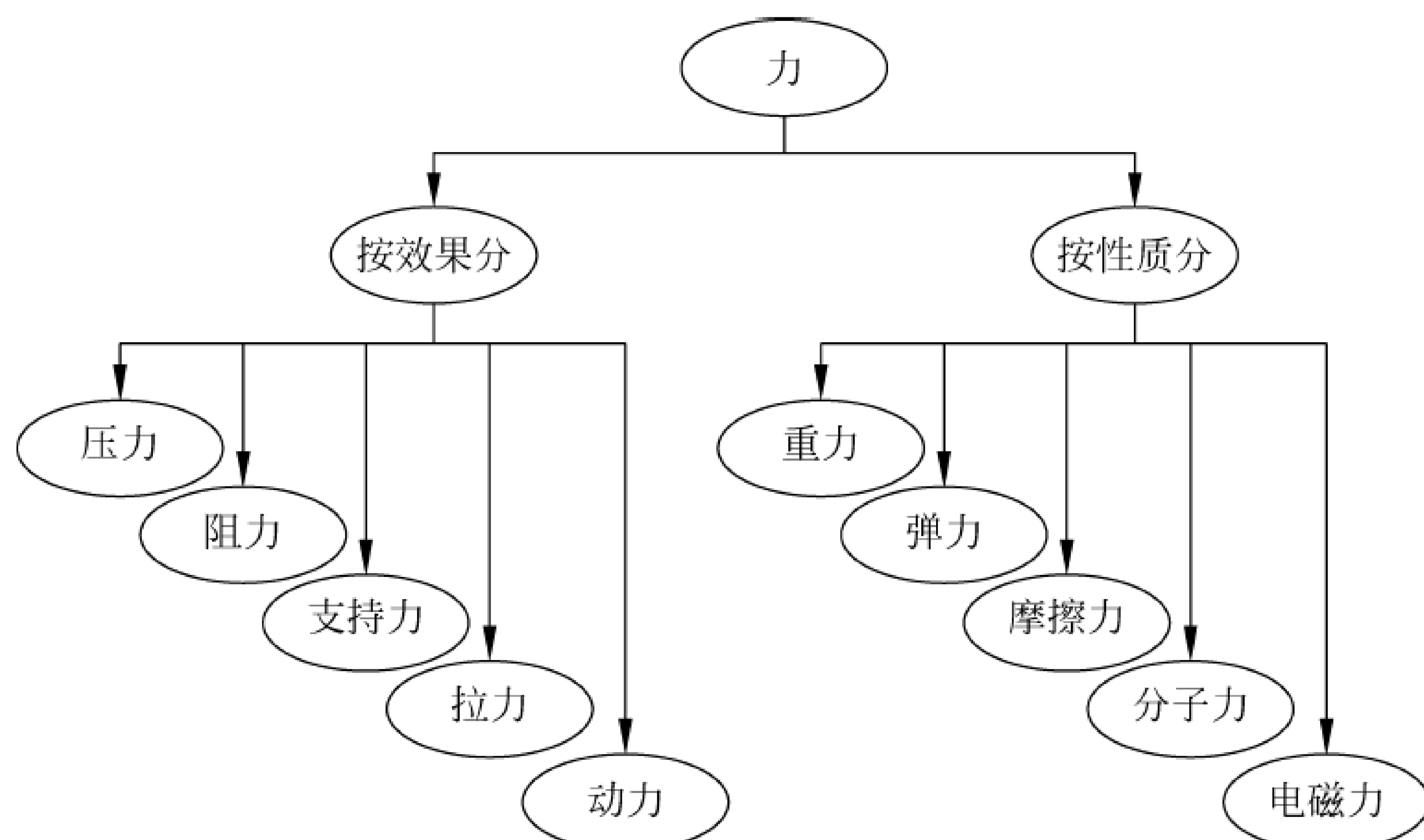



图 5-1-1 力的分类

**2. 根据力的概念分类**

如果单纯研究力的概念,我们主要认识场力(包括重力、万有引力、电场力和磁场力等)、弹力(包括弹簧的弹力、接触面的支持力和绳的拉力等)和摩擦力(包括动摩擦力、静摩擦力和最大静摩擦力等)等概念。其中场力的作用不需要物体接触,而其他力发生作用时,物体间必须互相接触。根据力的概念,其分类如图 5-1-2 所示。

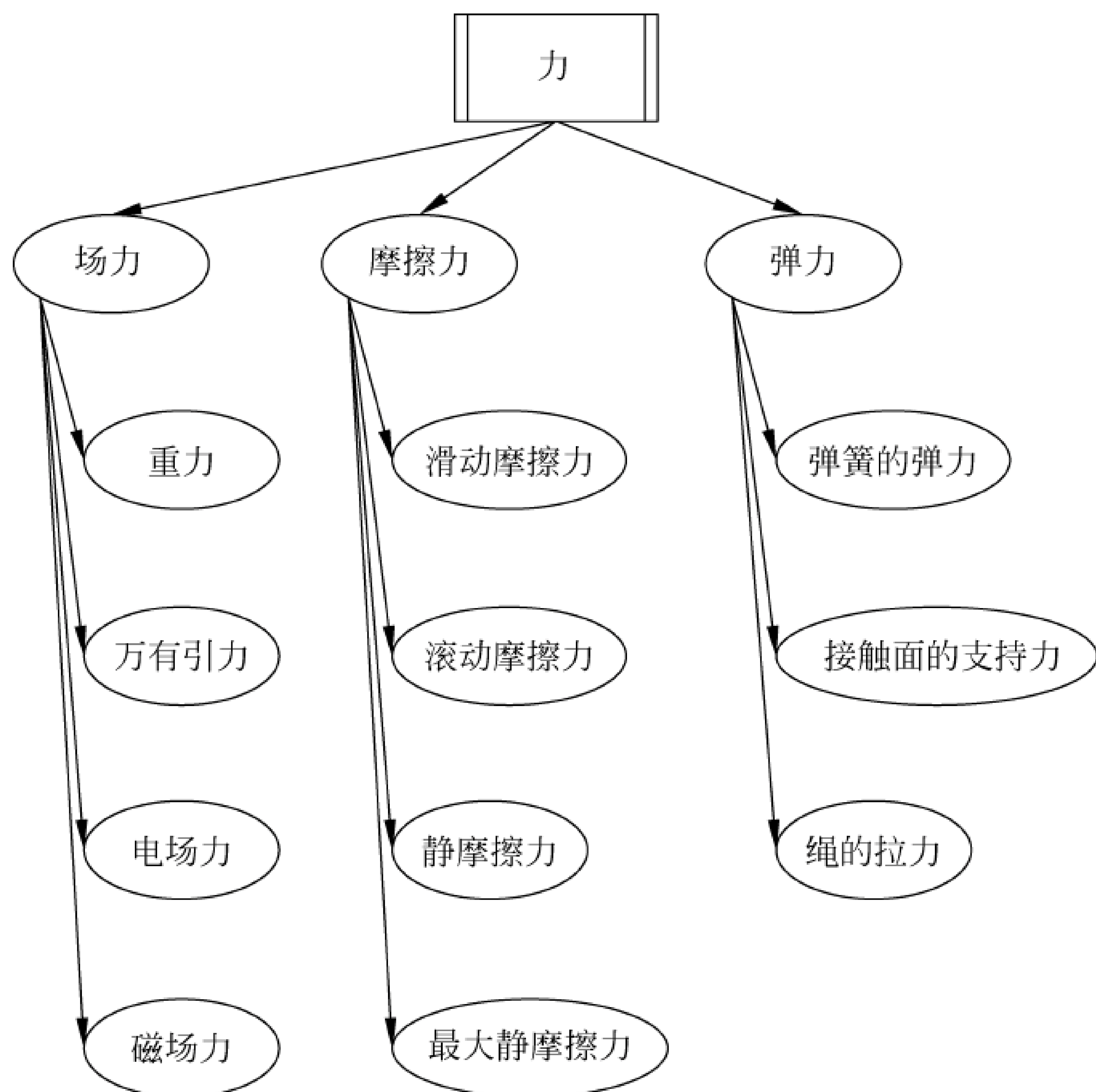



图 5-1-2 力的分类

### 3. 主动力和被动力

除此之外，为了便于力学研究，我们还将力分成主动力和被动力。重力、弹簧弹性力、静电力和洛伦兹力等有其“独立自主”的大小和方向，不受质点所受的其他力的影响，处于“主动”地位，称“主动力”。物体间的挤压力、绳内张力和摩擦力，它们没有自己独立自主的方向和大小，要看质点受到的主动力和运动状态，处于“被动地位”。具体分析如下。

1）主动力

（1）重力

重力：地球表面的质点受到地球的万有引力（视地球为惯性系）。

$$G=mg \tag{5-1-2}$$

式中 $m$ 为质点的质量，$g$ 为重力加速度，其大小与质点所处的纬度和高度有关，近似为 $9.8\mathrm{N/s^2}$，重力的方向为竖直向下。

（2）弹簧弹性力

弹簧水平放置如图 5-1-3(a)所示，一端固定，另一端与质点相连，处于自由伸展状态。以弹簧自由伸展时质点位置为坐标原点，沿弹簧轴线建立 $Ox$ 轴。沿 $Ox$ 轴拉动质点至某一位置，如图 5-1-3(b)所示，$x$ 表示质点坐标或对于原点的位移，$f_x$ 表示弹性力在轴上的投影，在弹性限度内，由胡克定律：弹簧弹性力的大小与物体相对于坐标原点的位移成正比：

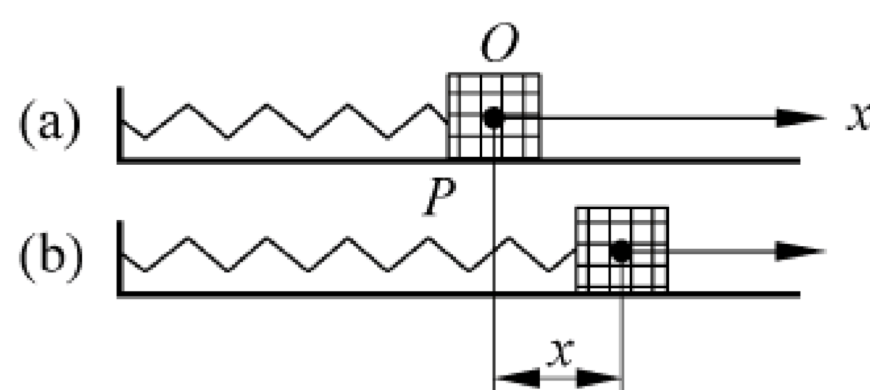


图 5-1-3　弹力的分析

$$\boldsymbol{f}=-k\boldsymbol{x} \tag{5-1-3}$$

式中负号表示弹簧弹性力的方向与位移相反，$k$ 是弹簧的劲度系数，与弹簧的匝数、直径、线径和材料等因素有关。

图 5-1-4(a)为被拉长的弹簧使小车向右运动，图 5-1-4(b)为被压缩的弹簧使小车向左运动。

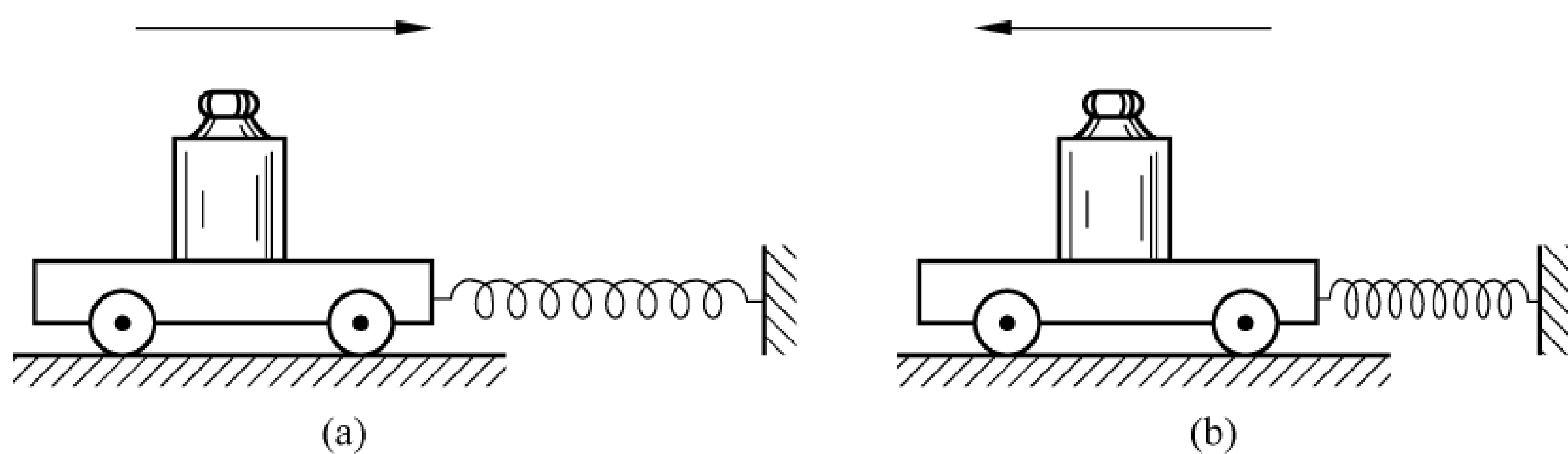


图 5-1-4　一些关于弹力的例子

（3）静电场力

① 库仑力：在真空中两个点电荷 $q_1$、$q_2$ 间的作用力跟它们的电荷量的乘积成正比，跟它们间距离 $r$ 的二次方成反比，作用力的方向在它们的连线上；同种电荷相互排斥，异种电荷相互吸引。

公式：
$$\boldsymbol{F}=k\frac{q_1q_2}{r^2}\boldsymbol{e}_r$$

② 一般静电场力：带电体周围存在电场，在电场内引入另一带电质点，则它受电场力的作用，这种力称为电场力。

$$\boldsymbol{F}=q\boldsymbol{E} \tag{5-1-4}$$

式中 $q$ 为质点所带电量,$\boldsymbol{E}$ 为质点所处位置之电场强度。上式表明:质点带正电时,$q>0$,$\boldsymbol{F}$,$\boldsymbol{E}$ 同方向;质点带负电时,$q<0$,$\boldsymbol{F}$,$\boldsymbol{E}$ 反方向。

(4) 洛伦兹力

有电流的空间存在磁场,磁场对运动着的带电质点有力的作用,这种力称为洛伦兹力。质点所受的磁场力为

$$\boldsymbol{f}=q\boldsymbol{v}\times\boldsymbol{B} \tag{5-1-5}$$

式中 $\boldsymbol{B}$ 为磁感应强度,$q$ 为质点所带电荷,$\boldsymbol{v}$ 为运动速度。上式表明:当质点所带电荷为正电荷时,$\boldsymbol{f}$、$\boldsymbol{v}$、$\boldsymbol{B}$ 的方向满足右手螺旋法则;当质点所带电荷为负电荷时,$\boldsymbol{f}$ 的方向与上述方向相反。

若质点既处于电场又处于磁场中,则运动带电质点所受的力为

$$\boldsymbol{F}=q\boldsymbol{E}+q\boldsymbol{v}\times\boldsymbol{B} \tag{5-1-6}$$

式(5-1-6)又称为洛伦兹公式(不是洛伦兹力公式 $\boldsymbol{f}=q\boldsymbol{v}\times\boldsymbol{B}$)。

2) 被动力(约束反作用力)

(1) 绳内的张力

在张紧绳索上某位置作与绳垂直的假想截面(图 5-1-5),将绳分成两段,这两段的相互作用力即该处绳的张力。处理类似问题时应注意,张力是由于绳索的拉伸形变而产生的,但形变量与原长相比很小,因此,绳的伸长量可忽略不计。

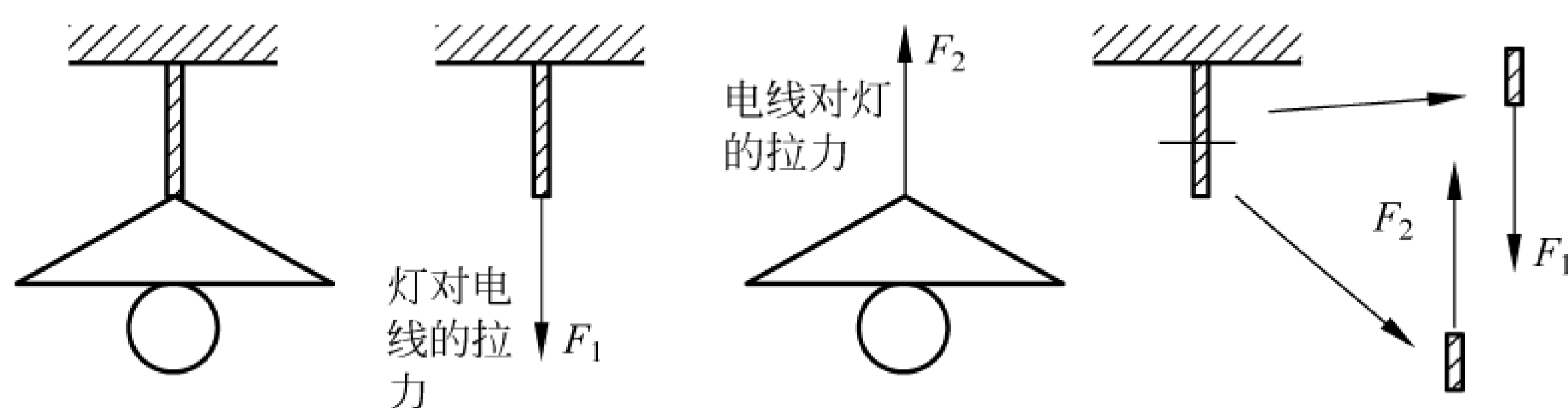



图 5-1-5 张力的分析

(2) 支持面的支撑力

两物体接触并压紧,双方均因挤压而形变,变形后的物体企图恢复原状而互相施以挤压弹性力(形变往往微乎其微,常忽略不计)如图 5-1-6 所示。

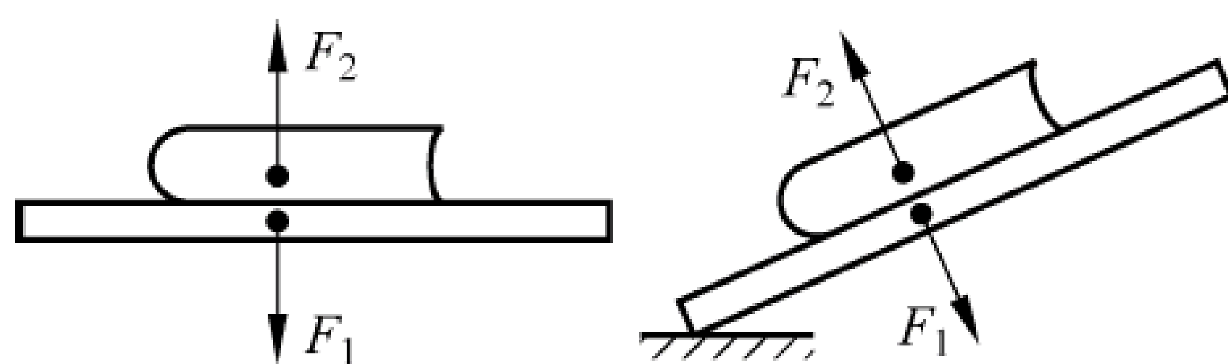



图 5-1-6 支撑力的分析

对于互相挤压的物体,可将相互作用力分为两个分力。一分力:沿接触面切线方向,另一分力:与接触面垂直,前者属于摩擦力,后者属于正压力。

(3) 摩擦力

固体间的摩擦力叫作干摩擦力,包含静摩擦力和滑动摩擦力。

① 静摩擦力 $f_0$,最大静摩擦力 $f_{0\max}$;

② 滑动摩擦力 $f$,$\mu$ 为滑动摩擦系数,$N$ 表示正压力,则有

$$f=\mu N,\quad f_0\leqslant f_{0\max}=\mu_0 N \tag{5-1-7}$$

其中:$\mu$ 和 $\mu_0$ 与物体材料、表面光滑程度、干湿程度及温度等多种因素有关,一般的计算中,视为常量,且 $\mu_0>\mu$。摩擦力的方向与相对运动或相对运动趋势的方向相反。

表 5-1-2 列出一些材料的摩擦系数。

表 5-1-2 一些材料的摩擦系数

| 材料(之间) | 滑动摩擦系数 | 材料(之间) | 滑动摩擦系数 |
|---|---|---|---|
| 钢～钢 | 0.25 | 钢～冰 | 0.02 |
| 木～木 | 0.30 | 木～冰 | 0.03 |
| 木～金属 | 0.20 | 橡皮轮胎～路面(干) | 0.71 |
| 皮革～铸铁 | 0.28 | | |

## 5.2 关于牛顿定律的话题

几百年来,人们从不同的角度理解牛顿定律,首先是对牛顿定律的不同表述形式;其次是对牛顿三个定律的逻辑关系的不正确理解,另外是对定律的物理内涵的不恰当扩展。

下面是对牛顿定律常见的表达方式。

牛顿第一定律表述1:一切物体总保持匀速直线运动状态或静止状态,直到有外力迫使它改变这种状态为止。

牛顿第一定律表述2:物体总保持原来运动状态,直至有外力迫使它改变这种状态为止。

牛顿第二定律:物体受到合外力的作用会产生加速度,加速度的方向和合外力的方向相同,同一物体,加速度的大小与合外力的大小成正比;同一合外力,加速度的大小与物体的惯性质量成反比。

牛顿第三定律:两个物体之间的作用力和反作用力,在同一条直线上,大小相等,方向相反。

然而,多数人并不十分清楚牛顿定律的真正逻辑关系,也没有深入思考牛顿定律隐含的丰富的物理哲学问题。例如,对牛顿定律做出的不恰当解释,认为牛顿第一定律只不过是牛顿第二定律的特殊情况;或认为牛顿第一定律不及其他两条定律重要;大部分教科书把牛顿第一定律说成是惯性定律;甚至有的学者还试图将牛顿三定律修改为两条定律。以上种种看法和意见各自都有自己的解释,但都没有注意到牛顿三定律的适用条件和范围等问题。

至今人们对牛顿定律体系的认识仍然维持在三条定律而不是两条定律基础之上,其根本原因就在于牛顿定律体系本身是完整的、相容的和严谨的,其蕴含着丰富的哲学思想和物理意义。本着探讨的目的,为了弄清其基本哲学思想和物理意义,深入理解好牛顿三大定律,本节将分别做较为细致的讨论。

### 5.2.1 牛顿第一定律

#### 1. 牛顿第一定律的建立

三百多年前,伽利略在通过结合实验方法与逻辑论证推翻了亚里士多德关于"力是维持物体运动的原因"的错误论断之后,通过大量的实验分析认识到,运动的物体受到的阻力越小,它的速度减小得就越慢,它的运动时间就越长。他还进一步通过推理得出,在理想情况下,如果表面绝对光滑,物体受到的阻力为零,它的速度将不会减慢,这时物体将以恒定不变的速度永远运动下去。

法国科学家笛卡儿(Rene Descartes,1596—1650)进一步补充了伽利略的结论,指出如果运动物体不受任何力的作用,不仅速度大小不变,而且运动方向也不变,将沿原来的方向匀速运动下去。

牛顿总结了伽利略等人的研究成果,从而概括出一条重要的物理定律:一切物体在没有受到外力作用的时候,总保持静止状态或匀速直线运动状态。这就是著名的牛顿第一定律。

牛顿第一定律是在大量经验事实的基础上,通过进一步的推理而概括出来的。我们周围的物体,都要受到这个力或那个力的作用,因此不可能用实验来直接证明这一定律。但是,从定律得出的一切推论,都经受住了实践的检验,因此,牛顿第一定律已成为大家公认的力学基本定律之一。

**2. 牛顿第一定律在经典力学理论系统中的重要地位**

1) 经典力学理论系统中的基本原理

首先我们可以将经典力学理论做如下简短归纳:

(1) 力与运动的关系定律——牛顿第一定律和第二定律;

(2) 普遍力的规律——牛顿第三定律和力的叠加原理(或力的独立作用原理);

(3) 具体力的规律——万有引力定律、胡克定律和摩擦定律。

2) 牛顿第一定律是其他原理的前提和基础

第一定律中包含的基本概念,奠定了经典力学的概念基础,从而使它处于理论系统中第一个原理的前提地位。这表现在:

(1) 首次批驳了延续两千多年的亚里士多德等人错误的力的概念,为确立正确的力的概念扫清了道路。

(2) 第一次科学地给出了力的定性定义(含力的本质和力的效果)。

(3) 第一次提出了经典力学的基本概念,为力学其他原理奠定了概念基础。

这些原理主要有:惯性参考系、力的定性概念(本质和效果)、惯性及惯性运动。牛顿第一定律首先表明了“不受力”下,物体的动力学规律(作惯性运动)。

在以上基础上才有牛顿第二定律,才有力学其他原理,为什么这么理解呢?

我们知道,经典力学的全部规律均要求在惯性参考系内考察,牛顿第二定律进一步要表明“受力”下,物体的动力学规律。在这里,利用力的效果——加速度 $a$,定量定义了力的大小和方向;也定量定义了惯性的大小即质量,得到牛顿第二定律表达式 $F=ma$。也就是说只有在力定性定义的概念基础上,才可以定量定义力,才能进一步表述力本身的规律,才可以研究反映作用力和反作用力的普遍关系的问题,即牛顿第三定律;才可以研究力的矢量性,进而研究叠加原理或力的独立作用原理,才能弄清力本身普遍的规律。

在此基础上进一步就可以研究具体力的规律,比如,万有引力定律、胡克定律和摩擦定律等。

总之,在力学理论系统中基本原理的关系上,无论从概念的关系上看,还是从表述规律的内容上看,第一定律均处于前提和基础地位。这在物理学发展的时间顺序上也一样。

**3. 牛顿第一定律的实验基础——“对接斜面”的理想实验**

其实验过程这里从略,下面只谈谈伽利略的科学方法的意义。

伽利略第一次从实验出发,运用科学推理推翻了延续两千多年的错误的力的概念,这在物理学研究方法上具有十分重要的意义。物理学的正确认识只能来源于科学实验。也只有科学实验对错误观点的批判才是最强有力的。

亚里士多德错误的力的概念之所以能延续两千多年,正是由于人们过分相信了“直觉”而没拿起“实验”这一科学研究的武器。

爱因斯坦对伽利略的工作给予极高的评价:“伽利略的发现以及他所应用的科学的推

理方法是人类思想史上最伟大的成就之一,标志着物理学的真正开端”。伽利略被人们誉为“实验科学之父”是当之无愧的。

**4. 关于牛顿第一定律实验验证的讨论**

(1) 不能用实验直接验证

因为不受力的实验只能是理想实验,是无任何实验误差的思维实验,严格来说“不受力”的条件真实实验不能满足,只能靠思维的逻辑推理去把握。从这一意义上说,任何受力趋于零的真实实验(例如气垫导轨实验)都不可能直接验证第一定律。

(2) 根本的方法是实验的间接验证

即用实验多次验证第一定律或它与其他原理共同的逻辑推论的正确性。

(3) 逻辑论证仅仅是辅助方法

逐步减小真实实验中物体在运动方向上的受力,令其逐步趋于零。利用外推法推论:如果受力等于零,则物体会作匀速直线运动。即牛顿第一定律的表述是可信的。

这种逻辑论证的实质不过是为第一定律又提供了几种理想实验的基础,因此不能作为实验检验的根本标准。

**5. 牛顿第一定律的质疑纠正:不是牛顿第二定律的特例**

牛顿在构成他的力学理论体系时,提出了机械运动的三个基本定律,其中把牛顿第一定律置于第一个原理的地位。牛顿是精通数学公理化体系的,他一定知道理论基础构成的“逻辑简单性”原则。如果第一定律是第二定律的逻辑推论,他不会将第一定律作为一个独立的原理提出。讨论如下。

第一,从物理上考虑,“受力”下物体运动的规律不能包括“不受力”下物体运动的规律。这可从两方面说明。

(1) 有人从第二定律的表达式合力推断,如果 $F=0$,则 $a=0$(静止或匀速直线运动),即为第一定律。

其错误在于:不是“不受力”,而是“受力”下合力为零。虽然与“不受力”时运动状态相同,但此时 $a=0$ 不是第一定律的惯性运动。因为惯性运动必有“不受力”的前提条件。

(2) 有人从物体受一个力的第二定律 $F=ma$,认为 $F=0$ 为物体“不受力”的数学表达,则 $a=0$ 物体作惯性运动。

其错误在于:第二定律是建立在物体“受力”下真实实验基础上的,因此由它不能逻辑地推断 $F\to 0$ 时理想极限情况下的运动,$F=0$(不受力)时,物体的加速度如何,只能靠理想实验说明。第一和第二定律有不同的实验基础,真实实验不能包括理想实验,这就是两个定律彼此独立的实验根据。因此,两个定律是物体“不受力”和“受力”下两个本质不同的运动规律。

一般地说,一般结论可以包括特殊结论(例如受力大小不同),但不能包括极限情况(例如“不受力”)。因为在极限情况下原有概念或规律会发生质变,其结果不能逻辑导出而只能靠实验说明。例如“平均速度”在 $\Delta t\to 0$ 时的极限产生“瞬时速度”概念,两者本质不同。

第二,无论在力学发展史上,还是在理论系统中,牛顿第一定律都是第二定律的前提和基础。

第一定律引入的基本概念为第二定律的建立奠定了基础,没有第一定律就没有第二定律。因为,由第一定律定义的惯性系概念是第二定律成立的参考系;第一定律给出力的定性定义(本质和效果)是第二定律中进一步力的定量定义的基础;第一定律给出的惯性概念

是第二定律中进一步定义惯性大小(即惯性质量)的基础。

第三,第一定律比第二定律适用范围更广泛。

在经典力学范围之外,第二定律及原有力的概念不再有效。但第一定律的推广表述“一个不受外界作用的孤立体系保持原来的运动状态”仍可适用。

**6. 广义牛顿第一定律**

(1) 力的真实性问题

力是物体间的相互作用,是物体发生形变或运动状态变化的原因。从力效应的真实性可以判定,力是真实的。物理学中有各种性质的力:引力、弹力、摩擦力、电磁力、分子力等,这些都是真实力。但也有非真实的力,例如惯性力、惯性离心力、科氏力等,因为这种“惯性力”没有产生力的真实效应。

为什么会出现“惯性力”这一非真实力概念,并且在一些人看来它是真实的,甚至也有人猜测:惯性力是一种引力。

真实力与非真实力之所以界限不清,易混淆,原因在于牛顿定律表述上的不严格。

牛顿第一定律的表述是这样的:物体总保持匀速直线运动或静止状态,直到有外力迫使它改变这种状态为止。即力是使物体产生加速度(改变这种状态)的原因。据此,很容易得出这样一个推论:力的存在是相对的,因为运动状态是相对的,一个物体相对某一参照物作匀速运动,似乎不受力的作用,同一物体相对另一参照物又作加速运动,似乎又受力的作用。于是人们开始怀疑力的真实性问题,力是真实的存在还是我们的虚幻想象?但是,力确实是不以人的意志为转移的客观实在。这说明牛顿第一定律的表述存在问题。

牛顿第一定律应该改述为:物体总保持原来运动状态,直至有外力迫使它改变这种状态为止。这种表述适用于任何参考系,因此,不妨把它叫作广义牛顿第一定律。

广义牛顿第一定律中的“原来运动状态”既可能是匀速运动或静止状态,也可能是某种加速运动状态,这取决于所选的参考系。物体的这种不受外力时具有保持原来匀速或加速运动状态的性质可叫作物体的“广义惯性”。而物体在不受外力时原来运动状态的加速度,在此可把它叫作“惯性加速度”。

根据广义牛顿第一定律,物体不受外力的作用既可能作匀速运动,也可能作加速运动;物体受外力的作用,既可能作加速运动也可能作匀速运动。物体受力或不受力时究竟作何种运动,取决于所选的参照物(系)。因此,“力是使物体产生加速度的原因”这一说法是错误的,这种说法正是产生非真实“惯性力”的理论基础。

牛顿定律,作为广义牛顿定律的特例,仅适用于特殊参考系即惯性参考系。

(2) 什么是惯性参考系

现在的物理教科书中都是这样定义惯性参考系的:惯性参考系就是牛顿定律成立的参考系。仔细分析不难发现,虽然这种定义中所反映的牛顿定律与惯性系间的关系是正确的,但是,这一定义从逻辑上讲是循环定义,并不能说明什么是惯性参考系。本来,牛顿定律是在惯性参考系中的实验定律,什么是惯性参考系?惯性参考系就是牛顿定律成立的参考系,这种循环定义非但不能说明惯性系是什么,反而还会产生这样的歧义:非惯性系也是惯性系,因为在非惯性系中应用牛顿定律时设想了一个“惯性力”后,有些人把它又视为真实力,则牛顿定律又能成立,在这些人看来,非惯性系又是惯性系。

根据广义牛顿定律,我们可以这样来定义:惯性参考系就是惯性加速度为零的参考系。在这样的参考系中,物体不受外力时的加速度为零,即保持匀速直线运动或静止状态,受外

力时有下式成立：$F=ma$。也就是说，判断一个参考系是不是惯性系，要依据实验、观察或设想，看一个物体在这个参考系里不受外力时（或设想不受外力时）的加速度是否为零，当 $a_0=0$，则这个参考系是这个物体的惯性参考系；如果 $a_0$ 不等于零，则这个参考系不是惯性参考系。

（3）结论

总之，力学中应该严格区分真实力与非真实力。为此，可用“惯性加速度”取代“惯性力”这一概念，将牛顿第一定律改进为广义牛顿第一定律。

广义牛顿第一定律表述为：物体总保持原来运动状态，直至有外力迫使它改变这种状态为止。

原来运动状态的加速度叫“惯性加速度”。

### 5.2.2　牛顿第一定律与牛顿第二定律的关系

**1. 惯性的意义**

惯性是经典力学中的一个基本概念，同时它又是人们日常生活中的一个基础性观念，并且惯性问题也是经常被物理学界讨论的一个话题。可是，尽管经典力学经过了漫长的发展时期，人们在此问题上的讨论还比较混乱。

我们知道，惯性是物体保持静止状态或匀速直线运动状态的性质。一个物体，只要不受外力作用，原来静止的就会一直静止下去，而原来运动的则会一直作匀速直线运动。这里的问题在于：惯性是否是物体的性质？依据牛顿第一运动定律，任何物体均具有惯性。因而，看来惯性不是被研究物体的性质，因为这一性质是一切物体所具有的，也就是说它与物体的个别特征无关。因而，惯性只能是存在的一个特征，是被研究对象周围的环境在此对象上的表现。换一句话说，它是存在于物体周围的一种条件，一种约束。

20世纪初，德国数学家诺特尔（Emmy Noether，1882—1935）证明了：空间平移对称性导致动量守恒，空间转动对称性导致角动量守恒，而时间均匀性导致能量守恒。事实上，物体的惯性是时间均匀性与空间对称性的必然结果，因而它与个别的特殊研究对象无关。惯性不是个别存在物的性质，个别存在物只是惯性的显现者，惯性的本质与个别存在物的特性无关。从而我们就不能用反映个别存在物性质的量（例如质量）来测量惯性。因为惯性作为存在的一种显现，并无大小可言，它只是存在之状态的表达。

惯性，这个我们通常认为是由物体内在因素决定的性质，其实是物体存在方式的一种条件性：“试取汽车为参考系统来研究‘当汽车急剧刹车的时候，车中乘客有向前倾倒的倾向’这个问题，在汽车急剧刹车前，相对于汽车而言，乘客是静止的，在汽车急剧刹车时，乘客突然向前倾，这就是说，以汽车为参考系统，乘客由静止而突然向前倾，并不保持其静止状态，并不表现出惯性。”这个条件就是：物体要表现出惯性，它必须处于惯性参考系中。而事物的存在顽强地延续维持不变，无论运动是快是慢抑或停止，也只在惯性系中才成立。在研究物体的运动学与动力学问题时，惯性系总有着特殊的地位。

**2. 牛顿第一定律不是惯性定律**

人们对牛顿第一定律的一般理解是，物体都有维持静止或作匀速直线运动的趋势，因此物体的运动状态是由它的运动速度决定的，没有外力，它的运动状态是不会改变的。物体的这种性质称为惯性，所以，牛顿第一定律也被称为惯性定律。这样理解实际上只是领会了牛顿第一定律前半段的物理意义——惯性的含义，牛顿第一定律后半段阐述了力的本质，是关于力的定义，因此，把牛顿第一定律称为惯性定律是片面的。

**3. 牛顿第一定律与牛顿第二定律的关系**

当物体所受的合外力为零时,从牛顿第二定律可知物体处于静止状态或作匀速直线运动。可是,仅依据这一点却不能认为牛顿第一定律是牛顿第二定律的一个特例,因为这两个定律的论述对象其实是不一样的。牛顿第二定律的研究对象是一个具体物体,而牛顿第一定律论述的是整个存在的性质。惯性——这个任何物体均具有的性质其实不是我们的个别研究对象所具有的性质,因为这个“任何物体”,包括了天地间的万物,而万物的总称即是宇宙。也即任何个别的物体都不可能无条件地具有惯性;惯性是存在的特性,是存在着的时空的特性,是宇宙的特性。

其次,牛顿第二定律是关于个别物体因果性的规律,而牛顿第一定律却与个别物体的因果性无关,它是存在之状态的表述,它的表述是与具体的特定的时间无关的、瞬时性的。正是这种非时间性构成了牛顿力学的本质特征。也正是牛顿第一定律所成立的时间均匀性与空间对称性构成了惯性系的特殊地位,从而使我们可以在牛顿第二定律的意义上来研究物体的动力学关系。因为毫无疑问,物体的运动性质和规律与采用怎样的空间和时间来度量有着密切的关系。由此可见,不仅牛顿第一定律不是牛顿第二定律的特例,而且恰恰相反,牛顿第二定律正是牛顿第一定律所揭示的存在之状态在具体的个体事物上的展现。牛顿第一定律比牛顿第二定律具有更强的基础性。也就是说,正是惯性现象,构成了牛顿动力学成立的操作平台。由于物体在不受外力作用下保持其速度不变,因而物体运动速度的变化才跟物体的受力相关。

最后,牛顿把牛顿第一定律放在三个运动定律的首位也是与其对自然的信仰因素有关的。因为在文艺复兴之前的绝大部分思想家继承了亚里士多德关于物体运动内在决定论的观点。但在牛顿看来,基本的物质粒子完全是惰性的,没有任何自发的运动,而电、磁、光这些非物质的力量则成为神在自然中行动的载体。也就是说,牛顿第一定律内隐含着牛顿否定亚里士多德运动观的内在目的论从而建立新力学的形而上基础。

**4. 牛顿第一定律的表述方式**

牛顿第一定律是动力学定律的基础,但它本身并不表征物体的某种动力学性质,它是关于人类认知自然之美、自然之和谐的陈述。基于上面的论述,对牛顿第一定律的陈述方式作以下的要求是并不过分的:反映时间的均匀性,空间的对称性及自然之美对人的呈现。可是,现行的许多教科书中对牛顿第一定律的陈述是很不一致的。当然,这种不一致性用老眼光来看是无伤大雅的,但以今天的眼光来看,这种差异性就值得商讨了。

例如:一个物体,如果没有受到其他物体的作用,它就保持自己的静止状态或匀速直线运动状态。这样的陈述可能离惯性定律的本义较远,因为这一陈述的方式是在动力学的维度上来进行的,陈述的对象是“一个物体”。这和牛顿第二定律的研究对象是一致的,这样方式的陈述毫无疑问地可以把惯性定律认为是牛顿第二定律的一个特例,因为“如果没有”这几个字就表达了陈述事件的某种特殊性。

另外一种常见的陈述方式是:一切物体总保持匀速直线运动状态或静止状态,直到有外力迫使它改变这种状态为止。这样一种表述比前一种完整多了,它几乎就是牛顿的原义,但这里的“一切物体”应当换成“任何物体”。因为在此论述中的“任何物体”实际上是对一切物体的否定,而“有外力”应当换成“其他物体的作用”,因为惯性定律是不涉及力的,操作意义上的力这个动力学的基本概念与惯性无关。

我们试着这样来陈述牛顿第一定律:存在着的宇宙有这样一种性质,它使任何物体在

没有受到其他物体作用的时候总保持静止状态或匀速直线运动状态。或许，这样的一种陈述方式是较明晰的陈述方式，它强调了惯性与惯性的表现者(个别研究对象)的严格区分，这个陈述的主语是性质，这样的陈述才可称为关于“惯性”的定律。而我们也应当将惯性定义为：使物体保持静止或匀速直线运动状态的性质。但是，这样的陈述只是表达了牛顿第一定律一半内涵，即惯性的定义，这也是人们把牛顿第一定律称为“惯性”定律的原因，因而是不正确的，从牛顿三定律的整体来看也是不全面的。

### 5.2.3　牛顿第二定律

牛顿第二定律，主要内容是物体所受合力等于其质量和加速度的乘积，它是力学的最基本的定律，广泛应用于各种物理过程的分析。

**1. 牛顿第二定律的导出**

牛顿第二定律以牛顿第一定律为基础，定量地表述了力与物体运动状态变化的关系。主要有两种表述方式。

一种是通常的形式：物体运动量(后称动量)的变化率与合外力成正比。其数学公式为

$$\boldsymbol{f}=km\boldsymbol{a} \tag{5-2-1}$$

式(5-2-1)中 $\boldsymbol{f}$ 表示物体所受的力或合外力，$\boldsymbol{a}$ 为物体的加速度，$m$ 表示物体的质量。当物体受到的合力为零的时候，其加速度为零，则物体运动状态不变，此时物体静止或者作匀速直线运动。牛顿第二定律直接提供了物体所受合力与其加速度之间的关系。其所受的合力与加速度成正比，并且比例系数为物体的质量。从这个公式，我们可以直接得出，物体质量越大，其运动状态越难改变，物体质量越小，其运动状态越容易改变。

另外一种是牛顿提出的：

$$\boldsymbol{f}=k\,\frac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t} \tag{5-2-2}$$

式(5-2-2)中 $\boldsymbol{f}$ 表示物体所受的力或合外力，$m$ 和 $\boldsymbol{v}$ 分别表示物体的质量和速度，这个公式是式(5-2-1)的微分表达式。从这个表达式我们可以直接理解到，物体速度改变的快慢跟物体所受合力成正比，与物体本身质量为反比。$\frac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t}$表示动量 $m\boldsymbol{v}$ 的变化率，$k$ 则为比例常量。这个式子既适用于质量不变的情况，也适用于质量改变的情况，如火箭、导弹因不断喷射燃气而减少质量，或物体的运动速度接近光速时质量随速度增大而增加。

欧拉认为质量是不变的，即$\frac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t}=m\,\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=m\boldsymbol{a}$，即 $\boldsymbol{f}=km\boldsymbol{a}$，因此，进一步选取适当的单位，可得 $k=1$，$\boldsymbol{a}$ 即物体的加速度。

这些公式不仅适用于直线运动的物体，也适用于曲线运动的物体，例如，可以用这个公式推导出匀速圆周运动物体的向心加速度与所受合力之间的关系。

**2. 牛顿第二定律的表述**

牛顿第二定律用文字表达即为“物体的加速度跟作用力成正比，跟物体的质量成反比，加速度的方向和力的方向一致”。

在这个表达式中比例常量 $k$ 则随所选择的力、质量、加速度的单位而定。在国际单位制中，规定 $m=1$ 千克(kg)，$a=1$ 米/秒$^2$($\mathrm{m\cdot s^{-2}}$)时，$f$ 为 1 牛顿(N)，即 $k=1$，公式就简化为 $\boldsymbol{f}=m\boldsymbol{a}$。力 $\boldsymbol{F}$ 和加速度 $\boldsymbol{a}$ 都是矢量，因此上述公式一般应按矢量来运算，只在物体作直线运动时，$\boldsymbol{F}$ 和 $\boldsymbol{a}$ 也始终在这条直线上，可以按代数量运算。

### 3. 牛顿第二定律的意义

牛顿第二定律把质量和力不可分割地联系着,定量地为力下了定义。但牛顿在逻辑上的矛盾(但他自己没有感到很大的困惑)是:在把他的第二定律陈述为某些物理量之间的关系时,他本应该从这些物理量的定义下手的。但实际上他没有对不依赖于第二定律的质量与力二者下明确的定义。现在看来,最能避免在逻辑上陷入困境的办法,就是在实际上把牛顿第二定律当成一个下定义的关系表述:即牛顿第二定律是力的定量定义。

牛顿第二定律是动力学的核心定律,动量定理、动能定理以及机械能守恒定律等都是由它直接推导出来的。它也是自然科学中建立的第一个微分方程,它的应用就是微分方程的应用,在推动数学的发展上也起了重要作用。但是牛顿第二定律只适用于经典力学中,不适用于微观领域的量子力学。

## 5.2.4 牛顿第三定律

### 1. 牛顿第三定律的传统表述

牛顿第三定律是牛顿对力学基本原理的一个最具有独创性的贡献。在现有的教科书中,关于作用力的定义是:甲物体对乙物体有力的作用,乙物体也必对甲物体有力的作用,力的作用是互相的,它们相互以对方作为自己存在的前提,不能孤立地存在,我们把其中的任意一个力叫作作用力,另一个力叫反作用力。

牛顿第三定律涉及作用力和反作用力之间的关系,在1687年出版的《自然哲学的数学原理》中,牛顿指出:"每一个作用总是有一个相等的反作用和它相对抗;或者说,两物体彼此之间的相互作用永远相等,并且各自指向其对方。"很多教科书都这样表述:两个物体间的作用力和反作用力大小相等、方向相反,作用在一条直线上。而另一类教科书是这样表述:两物体相互作用时,物体1对物体2的作用力 $\boldsymbol{F}_{21}$ 与物体2对物体1的反作用力 $\boldsymbol{F}_{12}$,量值相等、方向相反,即 $\boldsymbol{F}_{21}=-\boldsymbol{F}_{12}$。

然而,是不是任何两个相互作用的物体其作用力与反作用力都大小相等方向相反,并在同一直线上呢?或者每一种作用都有一个相等的反作用呢?回答是否定的。

按现在观点,牛顿第三定律适用于接触力和超距力,即相互作用是通过场以无限速度传播的,引力也不例外,从宏观上说两带电运动质点之间的相互作用,除了电场力(库仑力)和万有引力外还存在磁相互作用力,即洛伦兹力(安培力),再没有其他相互作用了。

如果说两运动电荷(或电流)之间的相互作用不是直接的,而是间接通过电磁场(来传递)进行的—即运动电荷 $q_1\boldsymbol{v}_1$(或电流元 $I_1\mathrm{d}\boldsymbol{l}_1$)产生的电磁场给其场中的运动电荷 $q_2\boldsymbol{v}_2$(电流元 $I_2\mathrm{d}\boldsymbol{l}_2$)以作用,反过来 $q_2\boldsymbol{v}_2(I_2\mathrm{d}\boldsymbol{l}_2)$产生的电磁场又给其场中的 $q_1\boldsymbol{v}_1(I_1\mathrm{d}\boldsymbol{l}_1)$以反作用。

那么无任何理由要求此二力一定相等的话,这电磁场从哪里来的呢?不是外来场作用于带电质点上的,而是两带电质点产生的,正是由于两带电质点产生电磁场,才引起两质点发生相互作用的,否则不能产生相互作用。所以归根到底还是两个带电质点之间的相互作用。再说两质点之间的万有引力又如何呢,我们都知道宇宙星辰都在不断地运动着,且它们之间的相互作用也不是直接的,从现代观点来说,也是通过场来传递的。

然而无论带电质点运动还是静止,两个带电质点之间的相互作用的库仑力和万有引力都满足教科书上表述的牛顿第三定律的关系,但是,磁相互作用在一般情况下就不一定遵从牛顿第三定律,即使考虑总的相互作用,也就是上述三种力的合力,也不满足牛顿第三定律的关系。既然是任何两物体之间相互作用都满足作用力与反作用力之间的关系,或者每一

种作用都有一个相等的反作用，也或者两个物体间的相互作用总是相等的指向相反。为什么两运动电荷或电流元（也属两物体之间）间的相互作用就不满足此关系呢？

如果说这两个带电粒子（质点）间的电磁相互作用并不是在瞬间从一个粒子立即施加给另一粒子的，电磁相互作用是通过以有限速度（即光速传播的）传播引起的，那么两电荷间的电场力也是通过电场传播的，它们之间的作用力满足 $\boldsymbol{F}_{1e}=-\boldsymbol{F}_{2e}$ 的关系。

**2. 牛顿第三定律的适用范围**

1）牛顿第三定律适用的相互作用

（1）直接接触的物体间的相互作用

牛顿第三定律是建立在绝对时空以及与此相适应的超距作用基础上的。所谓超距作用是指，分离的物体间不需要任何介质，也不需要时间来传递它们之间的相互作用。也就是说相互作用以无穷大的速度传递。对于直接接触的物体之间的相互作用，不存在超距作用，牛顿第三定律严格成立。如物体间的弹力、摩擦力等接触力。

（2）引力相互作用与静电相互作用

对于相隔一定距离通过场（如引力场、电磁场等）以有限速度传播而产生的物体间的相互作用，会有时间的推迟，需要考虑推迟效应，如引力相互作用和电磁相互作用。其中引力相互作用，由于宏观物体的运动速度远小于引力场的传播速度（光速），且目前的实际观测并未发现推迟效应带来的影响，即可看作超距作用，因此，一般认为引力相互作用遵从牛顿第三定律。对于某些电磁相互作用，如静止电荷之间的静电相互作用，由于静电相互作用总是沿着两电荷的连线方向（称之为有心力），且电磁场的传播速度（光速）很大，对于静止的电荷可不计传递相互作用所需要的时间，推迟效应可以忽略，即静电相互作用可当作以“无穷大速度”传递的超距作用，因而牛顿第三定律对静电相互作用仍适用。可见，对于可看作超距作用的引力相互作用和静电相互作用，牛顿第三定律仍然成立。

2）牛顿第三定律不适用的相互作用

牛顿第三定律也有不适用的情况，其证明的过程留待读者思考。

（1）运动电荷间的相互作用

① 运动电荷间的磁场力

② 运动电荷间的电场力

（2）运动电荷与静止电荷间的相互作用

**3. 对牛顿第三定律表述的修改的讨论**

上文所讨论的牛顿第三定律的适用范围，是基于对其教科书上普遍的表述，一般来说并不严格。站在深入和发展的高度上，应对牛顿第三定律表述做出修改。

根据实例与理论的分析，牛顿第三定律应这样表述较为恰当些：

任何两个相互作用的物体，如果相互作用力在同一直线上（这是充分条件），那么作用力与反作用力必定大小相等方向相反（分别作用在这两个物体上）。

这样修改后，就不会出现宏观质点在惯性系中低速运动时，它们之间相互作用的矛盾了。

应说明的是上述条件是充分条件，而不是必要条件，因为反之不一定成立。也就是两个相互作用力大小相等方向相反时，它们不一定在同一直线上。

有的物理学者可能认为：牛顿第三定律不必进行这样的修改，因为带电运动粒子构成的系统中，它们之间还存在电磁场物质，两电荷之间并非直接相互作用的，而是通过电磁场

进行的,系统中是两电荷和电磁场三者共存的,就好像一根有质量的弹簧两端各连着物体运动一样,彼此之间通过弹簧发生作用的,所以把电磁场和实物(两带电粒子)间的相互作用及电磁场动量考虑进去,牛顿第三定律就成立了,故无理由要这样修改。

## 5.3 关于非惯性参考系

我们知道,牛顿定律只适用于惯性参考系。相对于惯性参考系作加速运动的参考系叫作非惯性参考系。

根据参考系相对于惯性参考系的加速度的不同,非惯性参考系可分为平动非惯性参考系和转动非惯性参考系两大类。例如,自转着的地球就是一个转动非惯性系,在研究大气环流这类大尺度的运动时,就不得不考虑非惯性因素的影响;身处加速运动的车厢内,应如何解释观察到的现象;生活在航天器中的宇航员要解释一些现象就不宜选取地球或太阳为参考系;另外,从非惯性系中去处理问题,也可提供另一条解决力学问题的途径,有时可能是更为简捷的途径。

为了使得在非惯性参考系中牛顿定律仍适用,我们可以认为在非惯性参考系中的一切物体都受到某种"惯性力"的作用,借以代替这一个参考系相对于惯性参考系的加速度。这样一来,在此非惯性参考系中仍可应用牛顿定律。

本节我们对平动非惯性参考系和转动非惯性参考系加以简单讨论。

### 5.3.1 平动非惯性参考系中的惯性力

若参考系相对于惯性系运动,固定于该参考系上直角坐标系的原点作变速直线运动,且各坐标轴的方向总保持不变,即为平动非惯性参考系。如图 5-3-1 所示,最初,系于小车天花板的小球在连线方向静止于小车内,因为小球在连线拉力和重力作用下处于平衡状态。

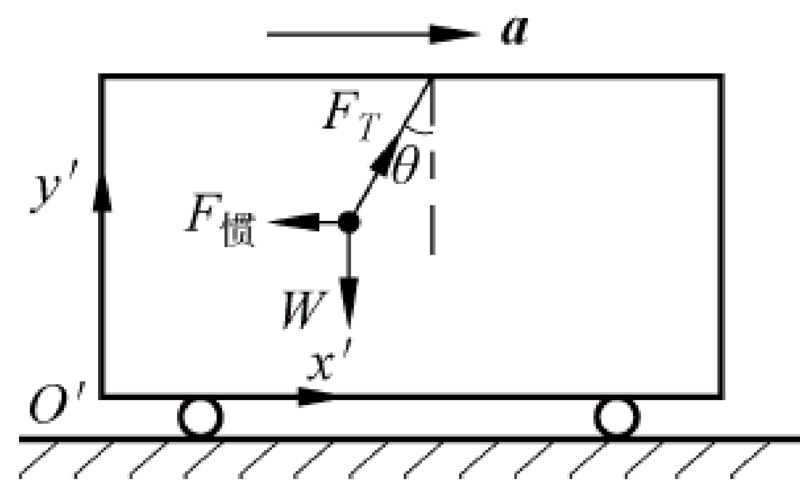



图 5-3-1 沿平直轨道以匀加速度 $a$ 行驶的车厢

然后,当小车相对于地面以加速度 $\boldsymbol{a}$ 作直线运动时,该车就成为一直线加速参考系。如果我们作为地面上观察者很容易理解,小球的连线不在竖直方向了,即这时小球所受连线拉力和重力不再是平衡力了,其合力使小球获得加速度 $\boldsymbol{a}$,从而使其与车身一起加速运动。然而,如果我们以小车为参考系来观察,这时小球所受连线拉力和重力不是平衡力,也就是小球所受的合力不等于零,但是,小球却与车厢保持静止状态!显然不符合牛顿第二定律。为了解决这个矛盾,我们以加速运动的车厢为参考系观察,可设想有一力 $\boldsymbol{F}_{惯}$ 作用于小球,其方向与小车相对地面的加速度 $\boldsymbol{a}$ 的方向相反,其大小等于小球质量 $m$ 与加速度 $\boldsymbol{a}$ 的乘积,即 $\boldsymbol{F}_{惯}=-m\boldsymbol{a}$,该力称为惯性力。这样,对于非惯性系,仍可沿用牛顿第二定律的形式,即小球相对于车身的加速度 $-\boldsymbol{a}$ 是惯性力 $\boldsymbol{F}_{惯}$ 作用的结果,这样,小球受到连线拉力 $\boldsymbol{F}_T$、重力 $W$ 和惯性力 $\boldsymbol{F}_{惯}$ 的作用,即小球所受的合力为零:$\boldsymbol{F}_T+\boldsymbol{W}+\boldsymbol{F}_{惯}=\boldsymbol{0}$,因此,与车厢保持静止状态。

在平动非惯性参考系中,质点所受惯性力 $\boldsymbol{F}_{惯}$ 与非惯性系的加速度 $\boldsymbol{a}$ 方向相反,且等于质点的质量 $m$ 与非惯性系加速度 $\boldsymbol{a}$ 的乘积,即

$$\boldsymbol{F}_{惯}=-m\boldsymbol{a} \tag{5-3-1}$$

惯性力和相互作用力不同。首先,惯性力不是相互作用,不存在惯性力的反作用力;还有,无论在惯性系还是在非惯性系,都能观测到相互作用力,但只有在非惯性系中才能观测

到惯性力。

## 5.3.2 转动非惯性参考系的惯性力

### 1. 离心惯性力

如图 5-3-2 所示，圆盘以匀角速度 $\omega$ 绕铅直轴转动，在圆盘上用长为 $\boldsymbol{r}$ 的轻线将质量为 $m$ 的小球系于盘心且小球相对于圆盘静止，即随盘一起作匀速圆周运动。从地面的惯性系观察，拉力 $\boldsymbol{T}$ 作为向心力，小球在 $\boldsymbol{T}$ 的作用下作匀速圆周运动，符合牛顿第二定律。如果以圆盘为参考系观察，小球受到拉力 $\boldsymbol{T}$ 的作用，但却保持静止，没有加速度，不符合牛顿第二定律。所以，相对于惯性系作匀速转动的参考系也是非惯性系，要在这种参考系中保持牛顿第二定律形式不变，我们引入惯性力 $\boldsymbol{F}_{惯}=-\boldsymbol{T}$，即

$$\boldsymbol{F}_{惯}=m\omega^2\boldsymbol{r} \tag{5-3-2}$$

称为惯性离心力，$\boldsymbol{r}$ 表示自转轴向质点所引矢量，与转轴垂直。

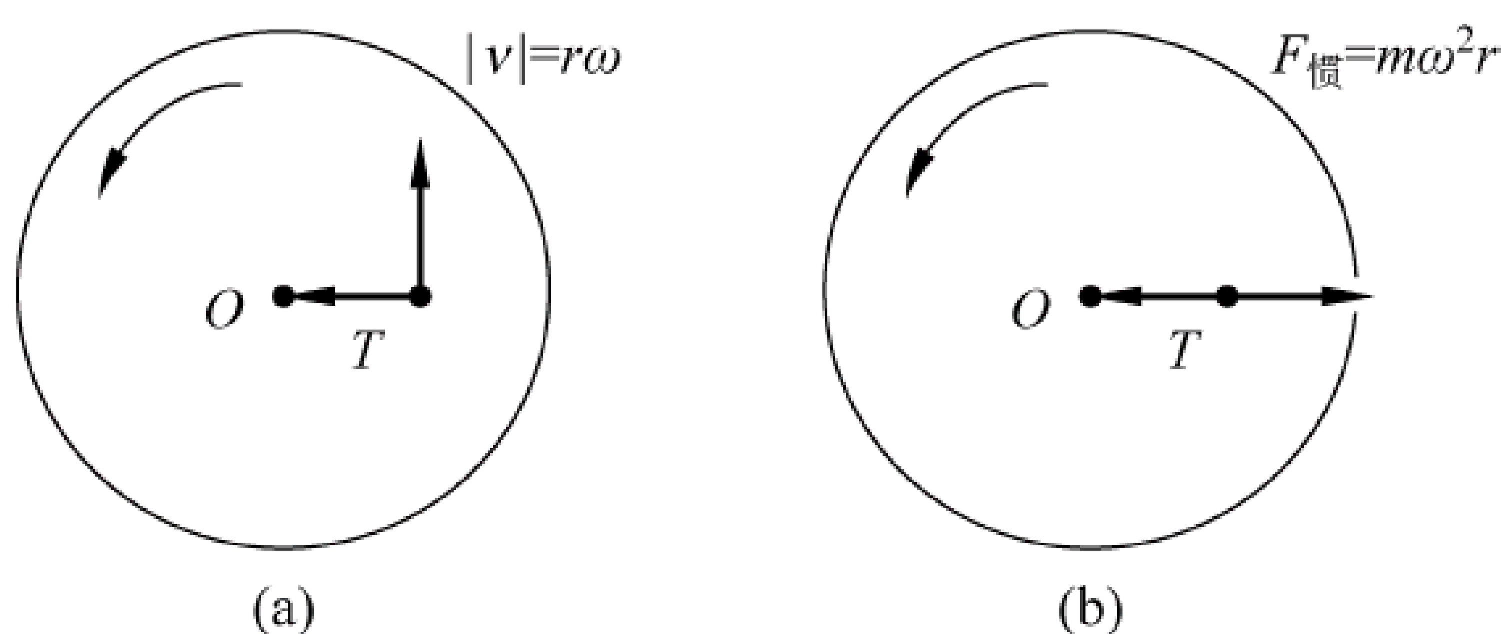



图 5-3-2 从惯性系(a)和转动非惯性系(b)解释小球的受力和运动

由此得出：若质点静止于匀速转动的非惯性参考系中，则作用于此物体所有相互作用力与离心惯性力的合力等于零，即

$$\sum\boldsymbol{T}+\boldsymbol{F}_{惯}=0 \tag{5-3-3}$$

在图 5-3-2 中，可认为小球所受线的张力与惯性离心力 $\boldsymbol{F}_{惯}=m\omega^2\boldsymbol{r}$ 平衡而静止。

### 2. 科里奥利力

若质点相对于转动的参考系运动，则质点还可能受到科里奥利力。如图 5-3-3 所示，一圆盘绕铅直轴一角速度 $\omega$ 转动。盘心有一光滑小孔，沿半径方向有光滑槽，其中置一小球 $m$，可视作质点，以细线连之。线另一端穿过小孔，可控制小球在槽中作匀速运动。现令小球以匀速 $v_{相}$ 沿槽向外运动。经很短时间 $\Delta t$，圆盘转过 $\omega\Delta t$ 角，而小球自 $A$ 运动至 $D'$。

现从地球惯性参考系研究小球的运动。在 $A$ 点小球具有径向速度 $v_{相}$，又有随盘转动而引起的切向速度 $\omega r_0$，$r_0$ 为 $A$ 点处半径。此二速度合成应使小球在 $\Delta t$ 时间内到达 $D$，但小球实际上到达 $D'$。这表明槽对小球作用有沿切线或圆弧方向的力。它使小球获得切向加速度，并使小球多走出弧长 $\overset{\frown}{DD'}$。显然，$\overset{\frown}{DD'}=(r_0+v_{相}\cdot\Delta t)\omega\Delta t-\omega r_0\Delta t$。因 $\Delta t$ 很短，可设小球以恒定加速度 $a_K$ 多走出 $\overset{\frown}{DD'}$，故 $\overset{\frown}{DD'}=\dfrac{1}{2}a_K(\Delta t)^2$，于是有 $a_K=2v_{相}\,\omega$。它应是由于槽壁作用于小球的推力产生的。用"角速度矢量"描述圆盘的转动，记作 $\boldsymbol{\omega}$。右手握拳并伸出拇指，四指指向圆盘旋转的方向，拇指即指向加速度矢量的方向，如图 5-3-3(a) 所示，加速度矢量的大小即角速率；角速度、$\boldsymbol{v}_{相}$ 和 $\boldsymbol{a}_K$ 形成右手螺旋，故

$$\boldsymbol{a}_K=2\boldsymbol{\omega}\times\boldsymbol{v}_{相} \tag{5-3-4}$$

式中 $\boldsymbol{a}_K$ 称作科里奥利加速度，是在惯性系中观察到的。槽施于小球的推力 $\boldsymbol{f}$ 的大小和方

向可用 $2m\boldsymbol{\omega}\times\boldsymbol{v}_{相}$ 表示,即 $\boldsymbol{f}=2m\boldsymbol{\omega}\times\boldsymbol{v}_{相}$。

再从圆盘非惯性系中观测,如图 5-3-3(b)所示,小球受到槽的侧向推力,但并未发生与槽垂直的运动,故必存在 $\boldsymbol{f}_{K}^{*}=-\boldsymbol{f}$,

$$\boldsymbol{f}_{K}^{*}=-2m\boldsymbol{\omega}\times\boldsymbol{v}_{相} \tag{5-3-5}$$

上式表示力与槽的侧向弹力平衡的关系,$f_{K}^{*}$ 称作科里奥利力或称科氏力。它同样不属于相互作用力的范畴,是在转动非惯性系中观测到的。它是科里奥利(Coriolis,Gustave Gaspard de,1792—1843)于 1835 年提出的。它和与位置有关的离心惯性力不同,因与相对速度有关,它不是场力,不能和重力场等相比。

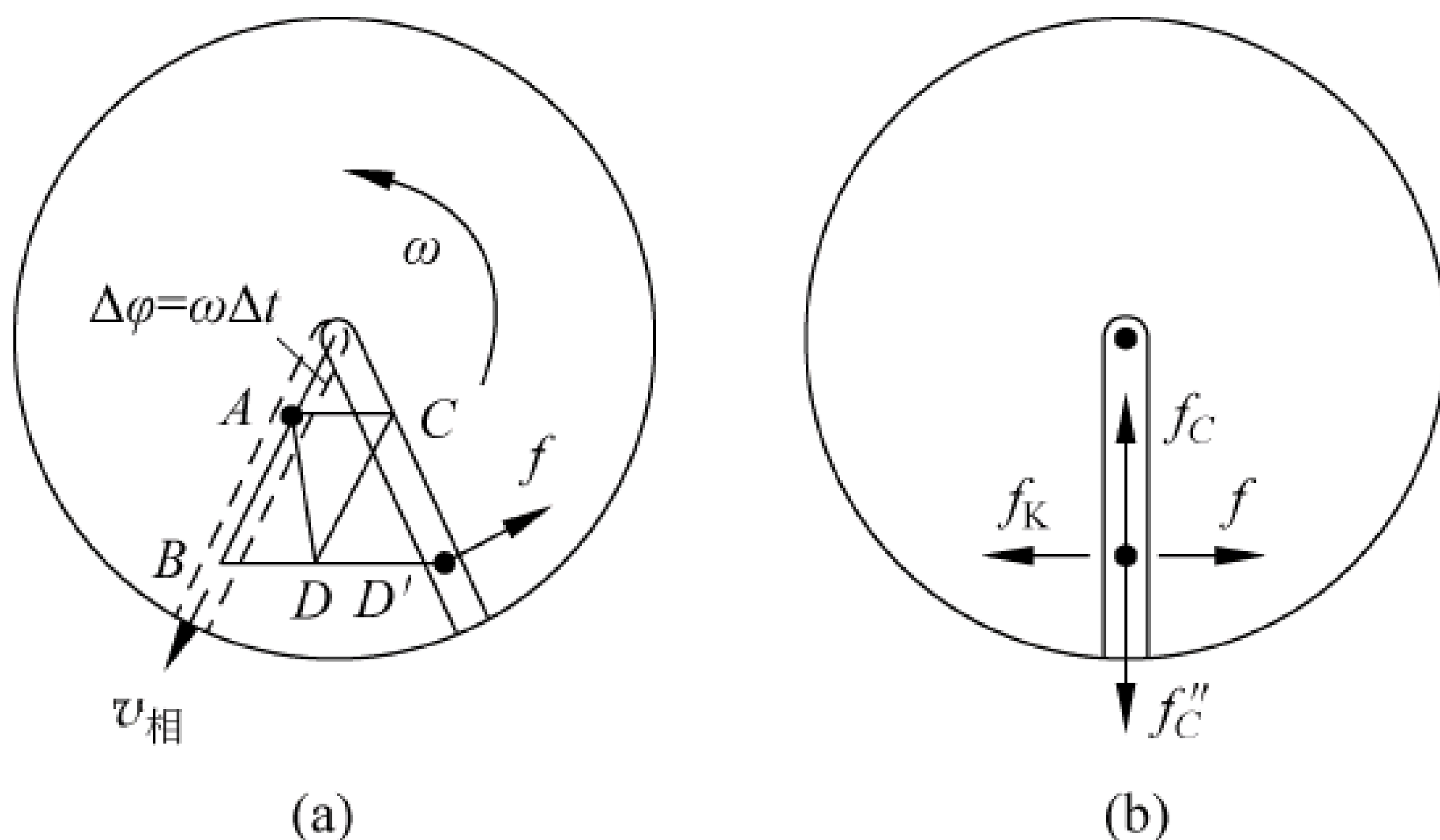



图 5-3-3

(a) 在惯性系看到 $f$ 力使小球走出附加位移;(b) 在非惯性系观察

以上就质点相对匀速转动参考系作匀速直线运动引入科里奥利力,但 $\boldsymbol{f}_{K}^{*}=2m\boldsymbol{\omega}\times\boldsymbol{v}_{相}$ 式也适用于转动参考系作变速转动和质点相对它作变速运动的情况。

### 5.3.3 地球上的科里奥利力学现象

由于地球是一个转动系,因此,凡在地面上运动的物体都要受到科里奥利力。在地球上我们可以观察到由科氏力导致的许多重要和有趣的力学现象。

**1. 傅科摆实验**

法国物理学家傅科(J. Foucault,1819—1868)于 1851 年在巴黎万神殿的拱顶上悬挂了一个摆长为 67m、摆锤质量为 28kg 的单摆,发现其摆平面作顺时针转动,32h 转一周,从而证明了地球的自转,这就是著名的傅科摆实验。由于摆绳很长,可以认为摆锤在水平面内运动。

**2. 河岸被冲刷**

在北半球运动的物体总受到从左侧指向右侧的力,由于科里奥利效应,河水的流向将向右偏转,加强了对右岸的冲撞。因此北半球河床的右岸受到比较严重的冲刷,右岸较陡峭。在我国黄河流域的河套地段,这一现象十分明显。而在南半球,情形与上述恰好相反,左岸比较陡峭。这里的“左”和“右”,是指观察者面向物体运动或河水流去的方向而言。

**3. 大气环流与涡旋**

由于地球自转,形成大气环流。赤道附近热空气上升,引起两极空气沿地面向赤道方向流动,这叫作信风。北半球信风自北向南,在科氏力作用下,这股向南的气流偏西,形成东北信风。而在南半球则形成东南信风,如图 5-3-4 所示。古代阿拉伯人凭借信风推动帆船出海通商,故亦将信风称为贸易风。

当地表某处遇热,热空气上升会形成低气压,外面的高气压空气要向低气压中心流动,

由于科里奥利力的作用，在北半球气流方向将向流动方向的右方偏转，从上向下看会形成逆时针方向的涡旋，如图 5-3-4 所示。夏季的天气预报气象云图中就常常出现这种涡旋式的强热带风暴图像。

**4. 落体偏东与箭、炮弹轨道的偏向**

高空自由下落的物体，其轨道总是向东偏移，这称为落体偏东现象。这个效应在北半球和南半球是一样的，在极地偏离为零，在赤道偏离最大。比如，如果物体在赤道从 100m 的高度下落到地面，东偏约为 2.2cm。

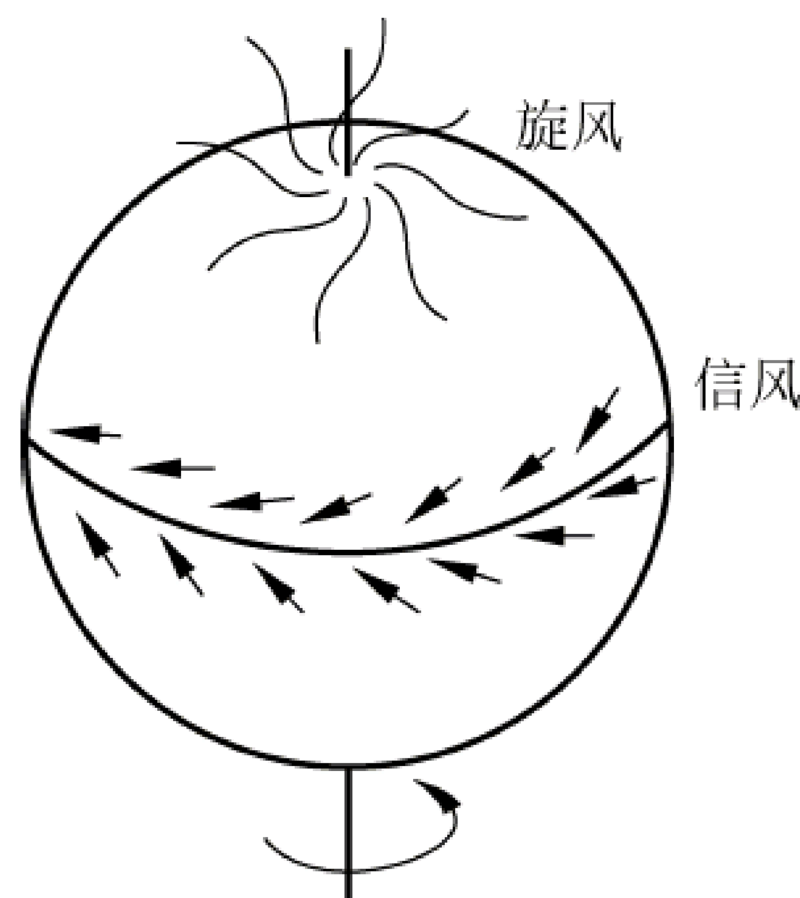


图 5-3-4　信风形成图

以地面参考系来计算洲际弹道导弹和人造地球卫星的轨道时，需要考虑科里奥利效应。例如，远程火箭发射升空，向上飞行时其轨道偏向西；向东水平飞行时其轨道偏向上；向西水平飞行时其轨道偏向下。

对于远射程的火炮，弹着点精度受科里奥利力的影响也不容忽视。20 世纪初，据说科技发达的英国已经将军舰火炮的瞄准器做了科里奥利力校正，但校正是以英国本土（北纬 50 度）为标准的。第一次世界大战期间英国与德国在马尔维纳斯群岛（南纬 50 度）发生海战，由于英国水兵科学素质有限，虽然他们努力瞄准，但由于双重的左偏误差使得炮弹经常落于德舰左侧而不能命中目标。

## 5.4　关于摩擦力的话题

摩擦力是大学物理学力学研究的三种不同性质的力之一，弹力的产生原因是弹性形变，其规律遵守胡克定律，实际上是一种理想情况；重力的产生是由于万有引力，其规律遵守万有引力定律；本节将回顾鲜为人知的人类对摩擦力产生原因和遵守规律的探索历史，探讨科学家在探索摩擦力奥秘中使用的思想与方法，介绍用于解释摩擦机理的凹凸说和粘合说，并讨论了关于生活中涉及的摩擦力应用的话题。

### 5.4.1　人类对摩擦力探索的历史

摩擦力是大家很熟悉的话题，但是提及人类对摩擦力认识的发展史，很多人一定很陌生。回顾摩擦的科学史，不但可以更清晰地理解摩擦力，而且也再现了人类思维的进步与技术的革新。

**1. 达·芬奇的研究**

对摩擦现象进行科学研究始于 15 世纪意大利的文艺复兴。文艺复兴产生了不少杰出的科学家和艺术家，其中有代表性的人物是雷纳德·达·芬奇（Leonardo Di Ser Piero Da Vinci，1452—1519）。人们只知道达·芬奇是天才的画家，却对他的科学成就缺乏了解。达·芬奇受到当时繁荣起来的造船技术的影响，开始着手于摩擦的实验，并将有关摩擦的见解散记于笔记中。1508 年，由于当时金属很贵，加工也不易，造船和建筑用的结构材料也主要以非金属为主，因此，达·芬奇使用石头和木头来做摩擦实验，并通过实验指出：物体的材料性质不同，摩擦力的大小也不一样。看上去这是一个很普通而当然的规律，但是有关论述摩擦力大小的历史，它却是最早的记述。达·芬奇对表面光滑和粗糙的东西的摩擦进行了比较，指出摩擦的原因归根到底在于固体表面的粗糙程度，即凹凸程度，这成为达·芬奇设想的重要基础。这一思想，直到进入 20 世纪之前，都几乎没有受到任何人的怀疑。

达·芬奇从轻、重物体的摩擦力的对比实验中，得出了结论：“一切物体，刚要开始滑

动,便产生叫作摩擦力的阻力。此摩擦力之大小,在表面是光滑的平面与平面之间摩擦的时候,为其重量的四分之一。”这指出了摩擦力与重量即与垂直力成正比,换言之,是最早确立了比例常数即摩擦系数这个概念。达·芬奇在摩擦研究领域还留下了不少卓越的设想、发现和启示。他第一个弄清楚了只要在两个摩擦面之间介入一层薄薄的第三者物质,摩擦力便产生很大的变化,这正是构成今天润滑技术的早期设想。他利用比较圆的沙粒介于其中,发现了摩擦力有大幅度减低的现象,并指出:“沙粒直径越大,摩擦力减小得越厉害。”他把这个事实归结为沙粒在表面之间滚动的缘故,并把摩擦分为“滑动摩擦”和“滚动摩擦”,指出这两种摩擦是由不同的原因引起的。一直到进入19世纪滚珠轴承和滚动轴承广为应用之时,滚动摩擦小于滑动摩擦的这一事实,才为大多数人所认识。

达·芬奇还进一步深入到摩擦的微细结构中,针对互相摩擦的两个表面,根据光滑抑或粗糙而获得的组合,或者同样是两个面,根据硬和软而获得的组合等进行了实验;并且还对表面碰伤的问题,即今天的所谓磨损问题也留下了记述。

达·芬奇还把在摩擦面上引起的现象粗分为三种,而且指出直接影响磨损现象的原因是材料性质的问题,考虑到硬碰硬、硬碰软和软碰软等三种材料性质的组合而造成的摩擦,进一步在这些组合中编进“直接接触”和“间接接触”的概念。可以认为这一概念是建立在当时发达的研磨技术(利用研磨粉研磨建筑用大理石等材料)之上的,这个磨损概念成为在17、18世纪发展起来的磨损凹凸说的基础。他十分注重观察,看一看钢缆卷起和把钢缆伸直两种状态,哪种状态下挪动它需要的人多。观察结果是两者都需要同样多的人数。他就从这个现象中抓住了接触面积对摩擦力的影响问题,正确地指出:同等重量的物体的摩擦力与其接触面积无关。

自达·芬奇以后,法国物理学家阿蒙顿(Amontons Grillaume,1663—1705)是在实验室内对摩擦系统地进行研究的最早的人。他在研究接触面积对摩擦影响时,使用如图5-4-1所示的实验装置:用放在上面的砝码 $G$,给予相叠数层的 $A$、$C$ 平板之间以垂直力,测量它们之间的摩擦力。用 $B$ 拉住 $A$ 板,以 $P$ 的力将 $C$ 板向右拉。根据变动板的张数进行实验的结果,得出了 $P$ 和 $G$ 之比(亦即摩擦系数)与摩擦面的接触面积之大小无关的结论。

此后人们以一定的载荷,使用接触面积不同的木头和大理石做了实验,发现摩擦力仅与载荷成比例,而与接触面的尺寸无关。

法国数学家帕朗(Parent Antoine,1666—1716)的功绩是,他开始把摩擦力作为新概念的“力”编进静力学的体系,解答了把摩擦力考虑在内的物体的静态平衡的问题。今天高等学校课本中谈到的“摩擦角”的概念,乃是帕朗于1704年的论文中最早谈到的。即如图5-4-2所示,设力 $T$ 作用于放在水平面上的重为 $P$ 的静止物体而开始滑动时,则作用于物体的反力 $N(N=P)$ 和摩擦力 $F_s$ 的合力 $R$ 与 $N$ 形成角度 $\theta_s$。因为 $F_s=\mu_s P=\mu_s N$ 及 $F_s=N\tan\theta_s$,所以,

$$\mu_s=\tan\theta_s \tag{5-4-1}$$

这个 $\theta$ 就是摩擦角,以 $N$ 方向为轴,将 $R$ 旋转360°时,形成顶角为 $2\theta_s$ 的圆锥,称为摩擦圆锥。

摩擦圆锥的物理意义在于:假定摩擦不存在,在小于 $\theta_1$ 的角度 $\theta$ 角的方向上施加一个斜切的力 $R$,物体必定开始滑动;若存在摩擦,只要 $\theta$ 处在摩擦角 $\theta_s$ 以内,则不管施加多大的力,物体依然是不滑动的。更广义一点讲,不管多大的或者来自任何方向的力作用于物体,只要它是作用在摩擦圆锥以内(小于摩擦角的角度),物体便继续保持静止状态。这在静力学上,是一个很有意义的重要发现。

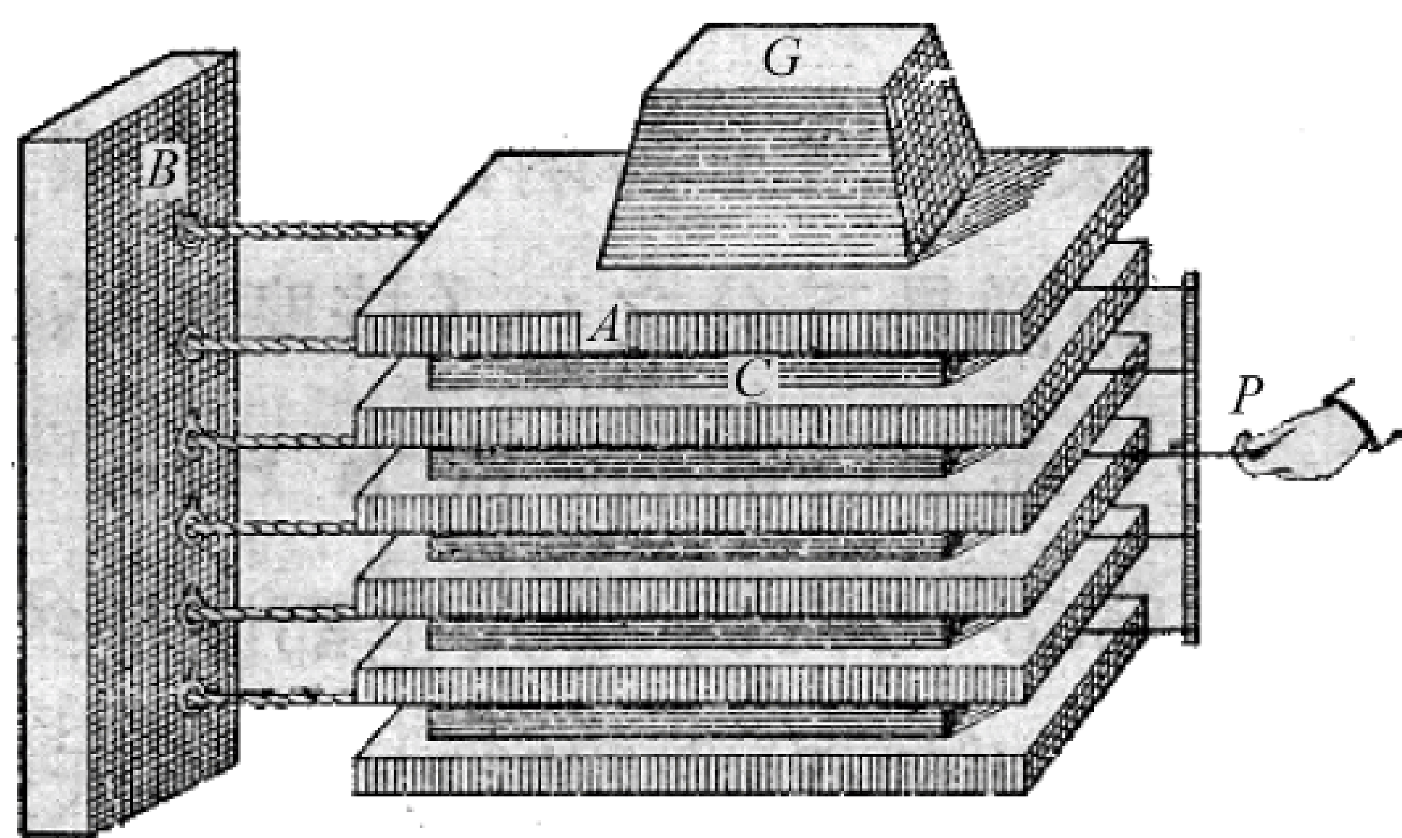



图 5-4-1 阿蒙顿的实验装置

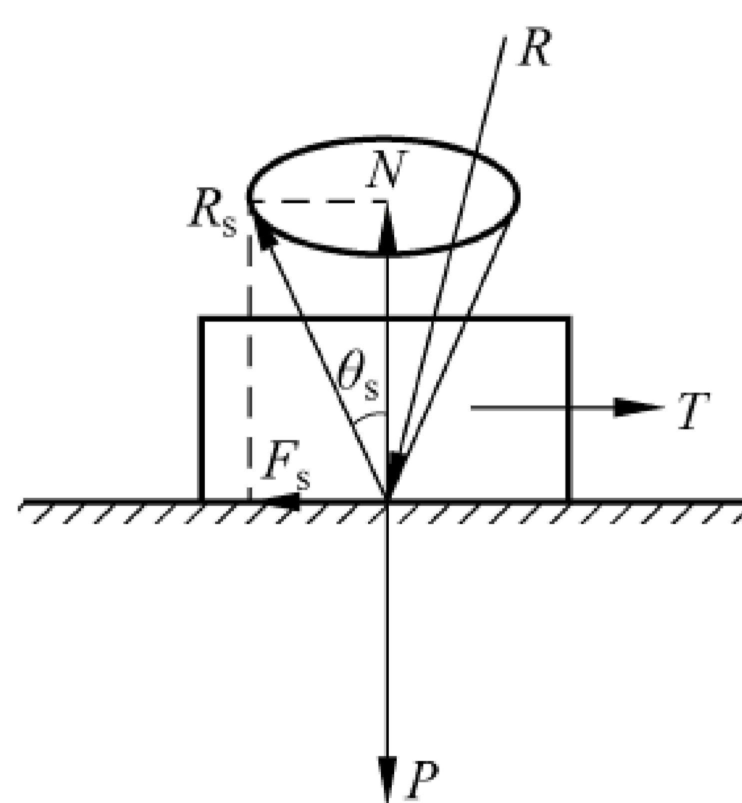



图 5-4-2 帕朗的摩擦角和摩擦圆锥

**2. 库仑的研究**

库仑的名字由于在静电学方面导出了著名的“库仑定律”而为人们所熟知，却不知在他的后半生，他在物理学和机械学的领域里确立了“库仑摩擦定律”，为探究摩擦力的规律做出一定贡献。当时，为了提高各种机械的效率和耐用性，摩擦成为设计工作上的难题；另一方面，之前所有摩擦实验均是在实验室内进行的，要应用于实际生产和技术的现场，数据的可靠性不高。因此，法国科学院悬赏征集有关摩擦的论文，库仑的论文是集 18 世纪摩擦实验和理论之大成，并得出了正确的结论。

库仑的工作，按照对象可以分为四类，第一是平面的摩擦，第二是绳缆的摩擦，第三是枢轴承的摩擦，第四是滚动摩擦。他充分了解各种复杂因素对摩擦均会产生影响，当物体在一个平面上开始滑动时，用实验来检查各种因素产生影响。接触面的材料性质与润滑状态、接触面的大小、加在接触面上的载荷以及接触之后到滑动前的时间。事实上他还考虑到大气的影响，特别是湿度的影响。湿度高，则水的分子附着于摩擦面上，就有起到润滑作用的可能性。但是根据他的实验结果，在实际上看不出湿度的影响，于是这个因素就被排除掉了。现在看来，这是不正确的。

库仑根据自己的实验求出了表示摩擦力和垂直力之间比例关系的式子：

$$F = A + \mu_1 P \tag{5-4-2}$$

式(5-4-2)中 $F$ 是摩擦力，$P$ 是垂直力，$\mu_1$ 是一种摩擦系数，与现在我们所熟悉的式子不同的是，他加进了 $A$ 这个常数。

在库仑的有关摩擦的研究中，他的另外一个贡献是对于滚动摩擦初次赋予了一个实验定律。摩擦分为滑动摩擦和滚动摩擦两种，这种情形早在雷纳德之后就有人指出过，由于滚动摩擦是由非常复杂多端的原因产生，目前还找不到能适合于所有场合的摩擦定律。

为了探究摩擦所遵守的规律，先辈科学家们在科学和技术进步的曲折道路进行了辛辛苦苦的研究。从这个艰难曲折的过程，我们可以深深体会到科学家们那种求实的态度和敢于探索的精神。所以有科学家认为，虽然摩擦力名不见经传，然而，人们摩擦规律探究的历史就是机械技术发展史的一部分。

## 5.4.2 摩擦的机理

**1. 物质表面的研究**

摩擦现象，归根结底是固体和液体的表面现象。固体表面可以从两个方面来观察研究。首先以宏观观点来看其外形，根据各种加工方法，既有镜面一般光滑的表面，也有锉子一般粗糙的表面。以我们现在所具有的加工技术来说，不能制成完全的几何平面，肯定存在凹

凸。以目前的加工技术来说，连最精细的加工表面，其凹凸高度仍有 $10^{-4}$mm 左右。

如图 5-4-3 所示，外形上的另一个重点就是，乍看很干净的固体表面，事实是由各种物质的薄膜覆盖着的，凹凸是金属表面的一般构造，在原来的金属基体上覆盖着三层或四层的异物质层，普通的脏污物质包括手指的油污或灰尘等各种脏污；吸附分子膜是来自大气层中的吸附层；加工变质层是指因车削或研磨而使金属的晶粒变得微细，通常成为比基体更硬的层。

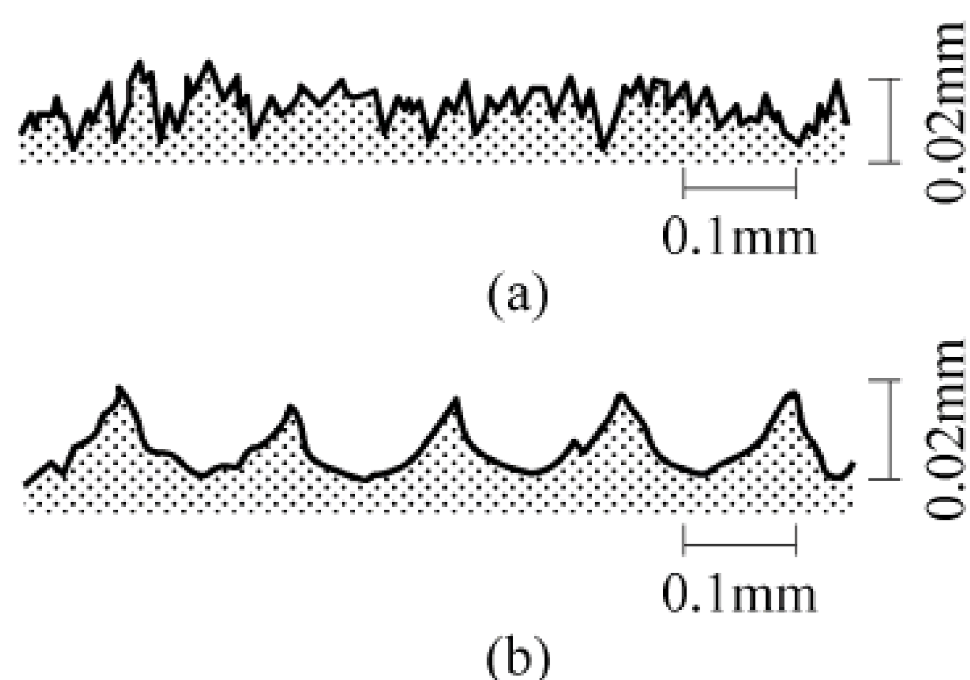


图 5-4-3　加工表面凹凸的放大图

与宏观的观察相反，还有微观的，即站在分子论立场上的观察。这就是从分子论的角度来看表面与内部究竟有什么不同。平常所见到的表面有液体和固体之分，两种固然有流动和固定的差别，但那只不过是构成两者分子和原子之间的分子间力和原子间力的大小不同而已，可以认为表面具有能量这一点在本质上是一样的。固体表面乍看是固定的，看上去好像是"死着的"，但却与液体表面一样，是"活着的"，即意味着表面具有能量，其最好的证明就是表面张力。从这个层面上分析，越光滑的表面，"摩擦力"会越大。

对于固体表面实在接触点的接触方式，存在着凹凸说和粘合说两种理论。

**2. 凹凸说**

凹凸说认为接触二面的凹凸部分主要是彼此以噬合的状态相接触。而且主要是在凸部上以互压在一起相接触而形成摩擦力。

不管进行何种表面加工、表面处理，在表面上或大或小地留下凹凸。这样两个面相压到一起时，由于两个面都不可避免地存在凹凸，就会出现若干复杂的噬合。实验测定表明，制作接触面积根据压力的大小，仅为外观接触面之数百分之一到数万分之一而已。

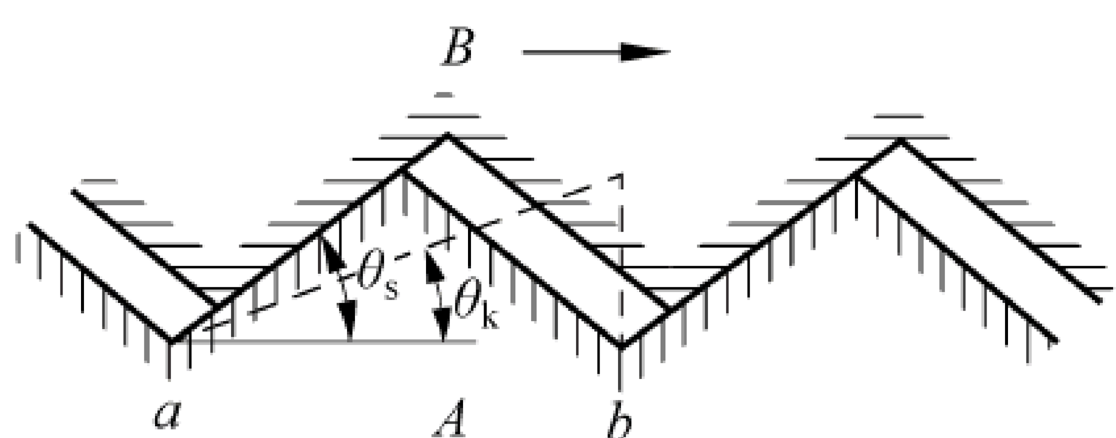


图 5-4-4　根据凹凸论的摩擦模

如图 5-4-4 所示，设 $A$、$B$ 二面的凹凸沿着斜面噬合。将 $A$ 固定，使 $B$ 朝着箭头方向滑动，则 $B$ 沿着斜面拱起来，因此相当于斜面摩擦，静摩擦系数：

$$\mu_s = \tan\theta_s \tag{5-4-3}$$

关于动摩擦系数 $\mu_k$，当 $B$ 水平地滑动于凹凸的一个间距 ab 之间，只是把 B 拱到凹凸二分之一高度，所以 $\tan\theta_k = \frac{\tan\theta_s}{2}$，因此：

$$\mu_k = \tan\theta_k = \frac{\mu_s}{2} \tag{5-4-4}$$

这两个式子是基于凹凸形状来解释摩擦起因的。

**3. 粘合说**

历史上，解释摩擦的起因还有粘合说。粘合说认为，在接触部分产生粘合现象是一个重要的前提，在固体表面的原子、分子的引力范围内，当对方的原子、分子接近或远离时，两者之间将出现能量损失。当固体表面非常干净、没有因污染产生的吸附膜时，表面处于高能状态，相接面的粘合就强，其剪切力也大，因而摩擦力就大。这一事实，成为支持粘合说的有力根据之一。

洗干净的玻璃表面的摩擦实验指出了凹凸说的不足。磨成达到凸透镜程度的光洁表面

和粗糙的加工表面做比较,原以为摩擦大致没有什么差别,可是实际上表明,充分研磨的表面摩擦力更大。而且摩擦的痕迹,最初宽 1μm 左右的伤痕,在摩擦过程中逐渐发展为大约 50μm 的程度。这个事实是正面否定凹凸说产生摩擦现象的有力证据。

在干净的表面,敷以极其多种多样的液体化合物薄膜,尽管不改变凹凸的情形,但证实了对摩擦系数的降低起着很大的作用。而霍尔姆在真空中做了非常干净的表面的摩擦实验,就连固体表面的一些微脏污,也将其置于真空中加热蒸发去除。这些实验结果说明,只要有微小吸附膜或因吸附膜的化学性质有了一点变化,就会造成比凹凸引起的大得多的摩擦系数变化,因此,粘合说最后当作摩擦的原理,被得到广泛承认。

摩擦现象的基本原理,作为物理学、化学、材料学、工程学的综合原理,而被归纳揭示出来。随着理论研究的逐步深入,凹凸说并没有作为误论而被废弃。事实上,基本现象即使是粘合,在因粘合而产生的摩擦当中,表面的凹凸大小也给予较大的影响。因此,凹凸说在进入 20 世纪时,在粘合说中得到了新的意义而恢复了生命力。

### 5.4.3　日常身边的摩擦现象

在我们周围的大部分固体的摩擦系数,是在 0.2～0.5 范围内,不管愿意不愿意,或者是在不知不觉中,我们的日常感觉和日常生活总是去适应这种现实的摩擦系数的大小。

步行时,如果鞋底和路面之间的摩擦系数减半,由于摩擦角减小,若仍按平常步行时的重心移动,无疑就会发生像初次滑雪和溜冰的人那样跌倒。又如果摩擦系数增大到二倍,就会像初次穿上田径赛用的钉子鞋一样,由于偶然的脚尖触碰也会绊倒。轮胎和路面或车轮和轨道之间的摩擦系数,是以 0.3 左右为基准而设计的,两者都是以引擎和马达的旋转力借轮胎或车轮的接触部位的摩擦传至路面或轨道,利用其反作用力来推向前进的,所以若把摩擦系数减半,哪怕继续增加动力,也像陷入泥泞中的轮胎那样,只会一味地打滑而不能随心所欲地开动车子。

摩擦作为日常生活的一种平凡的现象,几乎任何地方都存在,通常人们似乎已司空见惯。笔者某次开会曾经在某个经济相对发达地区的一个县级城市的步行街漫步,结果回来后,小腿特别难受,像下山之后的感觉(常言道:上山容易下山难!),原因是步行街街道的大理石打磨得太光滑!从摩擦力的发展史、摩擦的机理、各种各样的摩擦现象可以看到,摩擦力作为与我们生活息息相关的话题,深藏着有趣的奥妙。

## 5.5　自行车的受力分析

自行车自从 19 世纪诞生至今一百多年间,已发展成为一种轻便的交通和运输工具。它是目前结构最简单的交通工具,具有良好的运动性和有利于环保的特点。自行车在其车身设计上包括车架、车把、坐垫、脚踏板等,看来如此简单的机械能行驶得如此灵活,其中的奥妙充分体现了许多运动学、力学的原理,实现机械的最优化。下面将对自行车的运动进行力学上的分析,用以解释许多在自行车行驶时出现的现象。

### 5.5.1　自行车前进的动力

#### 1. 自行车前进时前、后轮的受力情况

如图 5-5-1 所示,设自行车在水平刚性的路面上向左行驶。前、后轮滚动时,轮子边缘上各点与地面接触的瞬间,与地面无相对滑动,也就是指轮子在作无滑滚动,其边缘在与地面接触时,相对于地面的瞬时速度为零,因此,前、后轮受到的摩擦力

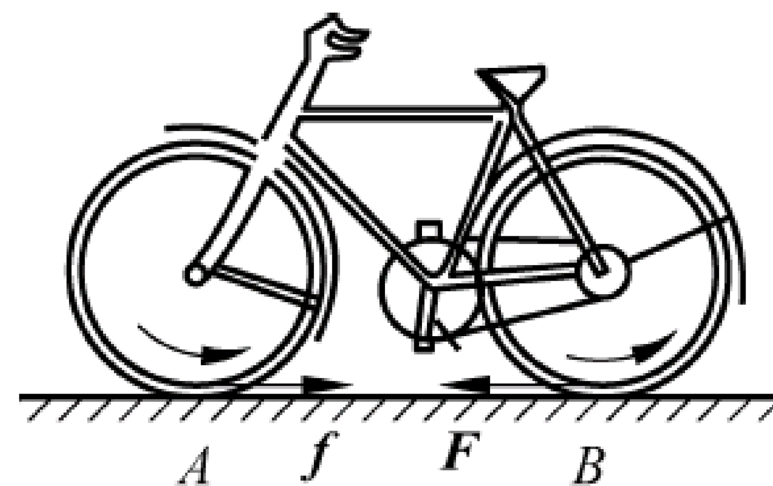



图 5-5-1　自行车前后轮受到的摩擦力

是静摩擦力。前轮是被动轮,在车身的推动下,前轮与地面接触点 $A$ 有向左滑动的趋势,则地面对前轮作用的静摩擦力 $\boldsymbol{f}$ 向右。正是这个静摩擦力使前轮转动,并因而作无滑动的滚动。后轮是驱动轮,它通过脚踏板做功而获得转动动能。后轮转动时,与地面的接触点 $B$ 有向右滑动的趋势,因此,地面对轮作用的静摩擦力 $\boldsymbol{F}$ 向左。这个向前的静摩擦力 $\boldsymbol{F}$ 就是自行车所受到的牵引力。自行车前进的动力来自于静摩擦力,如果没有这个牵引力,自行车根本不可能行驶。

**2. 自行车动能的来源**

自行车动能的增量完全来自内力所做的功,这个牵引力不做功,并没有为自行车提供动能。然而,牵引力的存在是自行车质心增加动能的必要条件。没有牵引力,内力所做的功无法转化为自行车的动力。下面是对这个问题的分析和说明。

如果将人和自行车看作一个质点系,同时,在这个问题的讨论上,忽略前轮受到的摩擦力及自行车受到的其他外界阻力,则水平方向受到的外力有牵引力 $\boldsymbol{F}$,脚蹬踏板的力及各部件之间的作用力,后两者均为内力,以 $\boldsymbol{F}_i^{(i)}$ 表示。由于车轮作无滑滚动,后轮与地面接触点的速度 $\boldsymbol{v}_B$ 为零,即在 $\mathrm{d}t$ 时间内的元位移 $\mathrm{d}\boldsymbol{r}_B=0$,故牵引力所做的元功 $\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}_B=0$,即总功也必为零。

根据质点系的动能定理,自行车动能的增量在数值上等于一切外力所做功与一切内力所做功的代数和。即

$$\mathrm{d}\left(\sum_{i=1}^{n}\frac{1}{2}m_i v_i^2\right)=\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}_B+\sum_{i=1}^{n}\boldsymbol{F}_i^{(i)}\cdot\mathrm{d}\boldsymbol{r}_i \tag{5-5-1}$$

其中,$m_i$,$\mathrm{d}\boldsymbol{r}_i$ 和 $v_i$ 分别是第 $i$ 个质点的质量、元位移和速率。由于外力做功为零,则上式可整理为

$$\mathrm{d}\left(\sum_{i=1}^{n}\frac{1}{2}m_i v_i^2\right)=\sum_{i=1}^{n}\boldsymbol{F}_i^{(i)}\cdot\mathrm{d}\boldsymbol{r}_i \tag{5-5-2}$$

可见,自行车动能的增量完全来自内力所做的功。

牵引力的存在是自行车质心增加动能的必要条件。

根据柯尼希定理,自行车对地面的动能等于质心 $C$ 的动能与各质点相对质心动能之和,故有

$$\mathrm{d}\left(\sum_{i=1}^{n}\frac{1}{2}m_i v_i^2\right)=\mathrm{d}\left(\frac{1}{2}Mv_C^2\right)+\mathrm{d}\left(\sum_{i=1}^{n}\frac{1}{2}m_i v_i'^2\right) \tag{5-5-3}$$

其中,$v_C$ 为质心的速率,$v_i'$为质点 $i$ 相对质心的速率,$M=\sum\limits_i m_i$ 为各质点质量之和。将式(5-5-3)代入式(5-5-2),得

$$\sum_{i=1}^{n}\boldsymbol{F}_i^{(i)}\cdot\mathrm{d}\boldsymbol{r}_i=\mathrm{d}\left(\frac{1}{2}Mv_C^2\right)+\mathrm{d}\left(\sum_{i=1}^{n}\frac{1}{2}m_i v_i'^2\right) \tag{5-5-4}$$

式(5-5-4)表明,内力所做的功等于质心动能的增量与各质点相对质心动能增量之和。如果忽略自行车传动装置摩擦力所做的功,则内力所做的功就是蹬脚踏板所做的功。由内力做功所获得的机械能,要有效地传递给质心,使质心增加动能,正是静摩擦力这个牵引力,为能量的传递提供了必要条件。

从质心运动定理 $M\boldsymbol{a}_C=\boldsymbol{F}$ 可看出,如果牵引力 $\boldsymbol{F}=0$,则自行车的质心加速度 $\boldsymbol{a}_C=0$。这种情况相当于自行车在光滑的冰面上时,无论车轮转动多快,车身并不向前运动,这就是

所谓的“打滑”现象，这时，由内力做功使后轮获得的机械能，不能有效地传递给质心。

从以上的分析可以知道，自行车受到的牵引力对它作用的质点 $B$ 不做功，因为 $B$ 对地面的绝对位移为零。但是，质点 $B$ 的绝对运动可看作它随质心 $C$ 的牵引运动和对质心 $C$ 相对运动的合运动。如图 5-5-2 所示，$\boldsymbol{r}_B=\boldsymbol{r}_C+\boldsymbol{r}'_B$。求微分得元位移的关系为 $\mathrm{d}\boldsymbol{r}_B=\mathrm{d}\boldsymbol{r}_C+\mathrm{d}\boldsymbol{r}'_B$，其中，$\mathrm{d}\boldsymbol{r}_B$、$\mathrm{d}\boldsymbol{r}_C$ 和 $\mathrm{d}\boldsymbol{r}'_B$ 分别为质点 $B$ 绝对元位移、牵连元位移和相对元位移，将上式代入式(5-5-3)，得

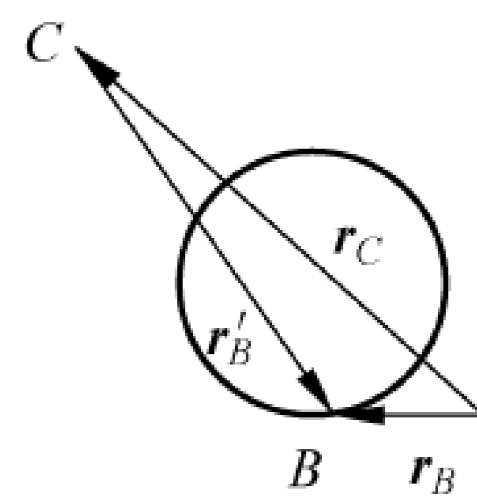



图 5-5-2　质点 $B$ 绝对运动的分解图

$$\mathrm{d}\left(\frac{1}{2}Mv_C^2\right)+\mathrm{d}\left(\sum_{i=1}^{n}\frac{1}{2}m_i v_i'^2\right)=\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}_C+\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}'_B+\sum_{i=1}^{n}\boldsymbol{F}_i^{(i)}\cdot\mathrm{d}\boldsymbol{r}_i \qquad (5\text{-}5\text{-}5)$$

由于 $\mathrm{d}\boldsymbol{r}_B=0$，所以，$\mathrm{d}\boldsymbol{r}_C+\mathrm{d}\boldsymbol{r}'_B=0$，则式(5-5-4)中有，$\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}_C+\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}'_B=0$，其中，$\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}_C$ 是牵引力对质点 $B$ 随质心的牵连元位移所做的元功，是正功；$\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}'_B$ 是牵引力对质点 $B$ 对质心的相对元位移所做的元功，是负功。这两个元功的大小是相等的。由此可见，牵引力对质点 $B$ 既做正功，也做负功。牵引力对质点 $B$ 做负功，使后轮的转动动能减少；同时，牵引力又对质点 $B$ 做等量的正功，使质心的平动动能增大。牵引力对质点 $B$ 做的总功为零。通过牵引力，实现了机械能的传递，这是关键。

## 5.5.2　自行车的动态稳定性

自行车的两个轮子和车架都是位于一个面内的，本身没有任何的支撑物。静止时不借助后脚架就无法竖直站立，但一旦运动起来就可以稳稳当当地行驶在路面上，并且自行车的速度越快，其稳定性表现得越强。即使车身有较大的倾斜，自行车也不会倒下，这里涉及刚体力学中回转效应运动的原理。

### 1. 回转效应的原理

在力学中，将旋转对称的刚体称为回转仪。陀螺就是其中的一个典型例子。当回转仪绕对称轴高速转动时，如果不受外力矩的作用，根据刚体角动量守恒定律，其动量矩保持不变，转动的角速度不变，即转轴的方向也不变。

当高速旋转的回转仪受到外力矩时，其自转轴就会绕另一个轴缓慢地转动，转动的方向总是力图使自转轴有向外力矩靠拢的趋势，这种转动称为进动。回转仪受到力矩作用产生进动的效应，就称为回转效应。如图 5-5-3 所示，陀螺在重力矩的作用下，一方面绕自转轴 $Oz$ 转动，另一方面自转轴又绕 $O\zeta$ 轴缓慢地进动，陀螺并不会倒下。

如图 5-5-4 所示，在 $t$ 时刻，回转仪的动量矩 $\boldsymbol{J}$ 的方向沿着角速度的方向，对于陀螺而言，受到的重力矩 $\boldsymbol{M}$ 与 $\boldsymbol{J}$ 方向垂直。根据动量矩定理，在 $\mathrm{d}t$ 时间内，这个力矩所引起的动量矩的改变量应该等于冲量矩，即

$$\mathrm{d}\boldsymbol{J}=\boldsymbol{M}\mathrm{d}t \qquad (5\text{-}5\text{-}6)$$

其方向与 $\boldsymbol{M}$ 一致。因此，在 $t+\mathrm{d}t$ 时刻，动量矩 $\boldsymbol{J}'$ 就如图 5-5-4 所示，为

$$\boldsymbol{J}'=\boldsymbol{J}+\mathrm{d}\boldsymbol{J} \qquad (5\text{-}5\text{-}7)$$

$\boldsymbol{J}'$ 的方向就是该时刻自转轴的方向。自转轴方向的这种改变就是进动，进动的方向是使 $\boldsymbol{J}$ 向 $\boldsymbol{M}$ 靠拢。如图 5-5-3 所示，重力矩保持与自转轴垂直，当自转轴变换到新的方向，重力矩也随之而变，自转轴又继续向重力矩靠拢，因此，回转仪不断地进动，而并不倒下。

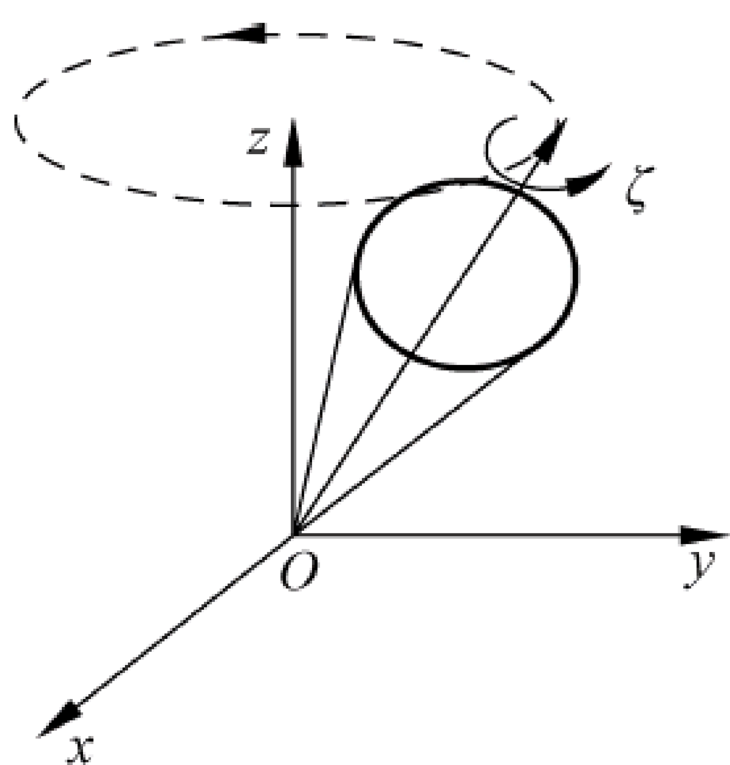



图 5-5-3 陀螺的进动

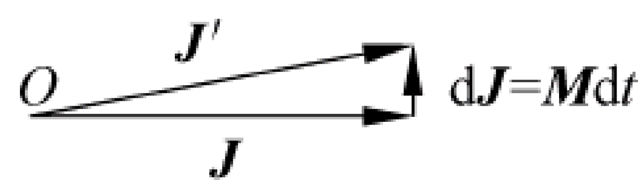



图 5-5-4 陀螺的角动量变化

**2. 自行车的回转效应**

自行车高速行驶时,前轮可视为回转仪。如果车身稍有倾斜,由于前轮的回转效应,车身并不倒下,这时前轮发生进动效应,进动方向使车轮的角速度的方向指向重力矩的方向,自行车就会转弯,随之出现反方向的惯性离心力,对倾斜给以校正,使车身恢复竖直,从而保证了自行车在行驶中的平稳。

### 5.5.3 自行车的极限速度

自行车的极限速度既取决于自行车的动力,更取决于自行车所受的阻力。自行车在运动过程中,其动力已在如上述所分析,其阻力主要是空气阻力,自行车上坡的时候还有坡度的阻力。

根据分析和研究,自行车的极限速度的计算公式为

$$v_{\max}=\sqrt{\frac{2(\tau_{驱}-\delta G)}{rC_D\rho S}} \tag{5-5-8}$$

其中,$S$ 为人的横截面积,$\rho$ 为空气密度,$C_D$ 为空气阻力系数,$\tau_{驱}$ 为驱动力偶矩,$G$ 为后轮所承受的重力,$\delta$ 表示原来的力与自行车轮轴中心 $O$ 点的垂直距离。

式(5-5-8)表明,空气阻力的影响很大,而滚动摩擦相对于空气阻力来说影响较小。在场地自行车比赛中,自行车车身尽可能轻,以减少滚动摩擦力偶矩的影响。

## 5.6 力学三个守恒定律及其成立的条件

在经典力学中,作为牛顿定律的推论,结合运动学规律可以推出三个定理,即动量定理、角动量定理及能量转换和守恒原理,而作为特例,合外力等于零、合外力力矩等于零和外力做的功与系统内非保守力做的功的总和等于零条件下的特例,可以得出相应的三条守恒定律。因此,其适用范围自然与牛顿定律相同,我们知道,在某些情况下,三条守恒定律比牛顿定律解决问题更方便,其物理含义也更加深刻。

在现代物理学中,三条守恒定律是作为物理学基本概念——时间、空间的"时空对称性原理"的逻辑推理的结论,不是从牛顿定律导出的。而时空对称性原理是基于大量实验事实提出的假说,是理论的出发点,其推论也还是要受实践检验。能量守恒定律对应时间流逝均匀性;动量守恒定律对应空间平移均匀性;角动量守恒定律对应空间旋转均匀性即空间的各向同性。因此,三条守恒定律适用范围远远大于牛顿定律;实践与理论证明,三条守恒定律不仅适用于宏观也适用于微观;不仅适用于低速也适用于高速。至今的一切实验结果都

验证了这一推论。正因为如此,在现代物理学中,动量、角动量及能量也是比“力”更为基本的概念。在整个物理学的理论系统中,三条守恒定律已经上升为原理。

### 5.6.1 动量守恒定律

**1. 动量守恒定律的导出方法**

(1) 结合牛顿第二定律、第三定律,容易导出动量守恒定律

设相互作用的 $A$、$B$ 两物体的速度分别为 $v_A$ 和 $v_B$,沿同一水平直线向同一方向运动,且 $v_A>v_B$,在 $A$、$B$ 两物体碰撞后各以 $v'_A$ 和 $v'_B$ 的速度仍沿原方向运动,求碰撞前后动量的变化。

根据牛顿第二定律得 $\boldsymbol{F}_A=m_A\boldsymbol{a}_A=m_A\left(\dfrac{\boldsymbol{v}'_A-\boldsymbol{v}_A}{t}\right)$,$\boldsymbol{F}_B=m_B\boldsymbol{a}_B=m_B\left(\dfrac{\boldsymbol{v}'_B-\boldsymbol{v}_B}{t}\right)$

根据牛顿第三定律有 $\boldsymbol{F}_A=-\boldsymbol{F}_B$,所以,$m_A\left(\dfrac{\boldsymbol{v}'_A-\boldsymbol{v}_A}{t}\right)=-m_B\left(\dfrac{\boldsymbol{v}'_B-\boldsymbol{v}_B}{t}\right)$,即

$$m_A\boldsymbol{v}_A-m_A\boldsymbol{v}'_A=m_B\boldsymbol{v}'_B-m_B\boldsymbol{v}_B \tag{5-6-1}$$

或者

$$m_A\boldsymbol{v}_A+m_B\boldsymbol{v}_B=m_A\boldsymbol{v}'_A+m_B\boldsymbol{v}'_B \tag{5-6-2}$$

式(5-6-1)表示,$A$、$B$ 两球碰撞后,$A$ 球动量减少量等于 $B$ 球动量的增加量。式(5-6-2)表示,$A$、$B$ 两球相互作用时,它们的动量之和保持不变。

(2) 从能量守恒和质量守恒定律出发,利用伽利略变换来导出动量守恒定律

在惯性参考系 $\Sigma$ 中,考虑两个自由粒子 1 和 2,它们的质量分别为 $m_1$、$m_2$,具有初始速度 $\boldsymbol{v}_{10}$ 和 $\boldsymbol{v}_{20}$。假设它们的初始位置及最终位置都相隔很远,这两个粒子在初始及最终阶段都没有相互作用,则它们的初始动能为 $\dfrac{1}{2}m_1v_{10}^2+\dfrac{1}{2}m_2v_{20}^2$。现让这两粒子碰撞,碰撞后的动能为 $\dfrac{1}{2}m_1v_1^2+\dfrac{1}{2}m_2v_2^2$,其中 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 是碰撞后的速度。由能量守恒定律得

$$\frac{1}{2}m_1v_{10}^2+\frac{1}{2}m_2v_{20}^2=\frac{1}{2}m_1v_1^2+\frac{1}{2}m_2v_2^2+\Delta\varepsilon \tag{5-6-3}$$

其中 $\Delta\varepsilon$(它可正可负)是由于碰撞所引起的粒子内部激发能的改变。内部激发可以是一种转动或者是一种内部的振动,也可以是束缚电子从低能态到高能态的激发。若为弹性碰撞,则 $\Delta\varepsilon=0$。这里,我们已假设粒子的质量 $m_1$、$m_2$ 在碰撞过程中保持不变。

现在从 $\Sigma'$ 系来观察这同一次碰撞,$\Sigma'$ 系以匀速 $v$ 相对于 $\Sigma$ 运动。在 $\Sigma'$ 系中,初速是 $\boldsymbol{v}_{10}$、$\boldsymbol{v}_{20}$,末速是 $\boldsymbol{v}_1$、$\boldsymbol{v}_2$。则有

$$\begin{cases}\boldsymbol{v}'_{10}=\boldsymbol{v}_{10}-\boldsymbol{v}, & \boldsymbol{v}'_{20}=\boldsymbol{v}_{20}-\boldsymbol{v}\\ \boldsymbol{v}'_1=\boldsymbol{v}_1-\boldsymbol{v}, & \boldsymbol{v}'_2=\boldsymbol{v}_2-\boldsymbol{v}\end{cases} \tag{5-6-4}$$

在 $\Sigma'$ 系中,能量守恒定律的表达式为

$$\frac{1}{2}m_1v'^2_{10}+\frac{1}{2}m_2v'^2_{20}=\frac{1}{2}m_1v'^2_1+\frac{1}{2}m_2v'^2_2+\Delta\varepsilon \tag{5-6-5}$$

我们假设参考系改变时激发能 $\Delta\varepsilon$ 不变,这是与实验相符的。

如果能量守恒定律在伽利略变换下具有不变性,那么,在 $\Sigma'$ 和 $\Sigma$ 两参考系中,初始动能都必须等于最终的动能加上内部激发能 $\Delta\varepsilon$。即是说,式(5-6-3)和(5-6-5)一定都成立。将

式(5-6-4)代入式(5-6-5),得 $\Sigma'$ 系的能量守恒定律为 $\frac{1}{2}m_1(v_{10}^2-2\boldsymbol{v}_{10}\cdot\boldsymbol{v}+v^2)+\frac{1}{2}m_2(v_{20}^2-2\boldsymbol{v}_{20}\cdot\boldsymbol{v}+v^2)=\frac{1}{2}m_1(v_1^2-2\boldsymbol{v}_1\cdot\boldsymbol{v}+v^2)+\frac{1}{2}m_2(v_2^2-2\boldsymbol{v}_2\cdot\boldsymbol{v}+v^2)+\Delta\varepsilon$。该式中如果

$$(m_1\boldsymbol{v}_{10}+m_2\boldsymbol{v}_{20})\cdot\boldsymbol{v}=(m_1\boldsymbol{v}_1+m_2\boldsymbol{v}_2)\cdot\boldsymbol{v} \tag{5-6-6}$$

那么这个表达式(5-6-6)就和在 $\Sigma$ 系中的能量守恒定律[式(5-6-3)]完全相同了。式(5-6-4)应对于任何 $v'$ 值都成立,因而有

$$(m_1\boldsymbol{v}_{10}+m_2\boldsymbol{v}_{20})=(m_1\boldsymbol{v}_1+m_2\boldsymbol{v}_2) \tag{5-6-7}$$

这正是动量守恒定律。

(3) 在经典力学中,一力学体系的状态可用拉格朗日量 $L(\boldsymbol{r}_i,\boldsymbol{v}_i,t)$ 来描述

当一保守体系的拉格朗日量在时空的某种变换下保持不变时,会导致一些守恒量。若假定空间是均匀的,则体系在平移(变换)下拉格朗日量 $L(\boldsymbol{r}_i,\boldsymbol{v}_i,t)$ 保持不变,即当体系各部分都平移 $\boldsymbol{\varepsilon}$ 时,当 $\boldsymbol{r}_i\to\boldsymbol{r}_i+\delta\boldsymbol{r}_i=\boldsymbol{r}_i+\boldsymbol{\varepsilon}$,相应地有 $L(\boldsymbol{r}_i,\boldsymbol{v}_i,t)\to L(\boldsymbol{r}_i+\boldsymbol{\varepsilon}_i,\boldsymbol{v}_i,t)=L(\boldsymbol{r}_i,\boldsymbol{v}_i,t)$,即 $\delta L=\sum\frac{\partial L}{\partial\boldsymbol{r}_i}\cdot\delta\boldsymbol{r}_i=\boldsymbol{\varepsilon}$。但是,$\delta L=L(\boldsymbol{r}_i+\boldsymbol{\varepsilon}_i,\boldsymbol{v}_i,t)-L(\boldsymbol{r}_i,\boldsymbol{v}_i,t)=0$。故,$\boldsymbol{\varepsilon}\cdot\sum\frac{\partial L}{\partial\boldsymbol{r}_i}=0$;由于 $\boldsymbol{\varepsilon}$ 是任意的,故 $\sum\frac{\partial L}{\partial\boldsymbol{r}_i}=0$,再利用拉格朗日方程 $\frac{\mathrm{d}}{\mathrm{d}t}\sum\frac{\partial L}{\partial\boldsymbol{v}_i}-\sum\frac{\partial L}{\partial\boldsymbol{r}_i}=0$,可得 $\frac{\mathrm{d}}{\mathrm{d}t}\sum\frac{\partial L}{\partial\boldsymbol{v}_i}=0$。因此体系的总动量

$$\boldsymbol{p}=\sum\frac{\partial L}{\partial\boldsymbol{v}_i}=\sum\boldsymbol{p}_i=\text{恒矢量} \tag{5-6-8}$$

即体系的总动量守恒。

**2. 动量守恒定律中部分守恒情况**

质点组的动量守恒定律包括全部守恒和部分守恒两种情况。

如果作用在质点组上的外力作用的矢量和为零时,那么质点组的动量亦即质心的动量将保持一恒矢量,若 $\sum\boldsymbol{F}_{i外}=0$,则可表为

$$\boldsymbol{p}=\sum m_i\boldsymbol{v}_i=m\boldsymbol{v}_c=\text{恒矢量} \tag{5-6-9}$$

其中 $m$ 为质点组的质量,即 $m=\sum m_i$,这是全部守恒情况。

如果作用在质点组上的外力在某一轴上的投影之和恒为零时,那么质点组的动量在该轴上的投影将保持一常数,若 $\sum\boldsymbol{F}_{i外}=0$,则

$$\boldsymbol{p}_x=\sum m_i\boldsymbol{v}_{ix}=m\boldsymbol{v}_{cx}=\text{恒矢量} \tag{5-6-9$'$}$$

这是部分守恒情况。

全部守恒情况的表述是式(5-6-9)和式(5-6-8),而部分守恒情况的表述式(5-6-9)′不具有普遍意义,因为如果 $x$ 轴相对惯性系是固定的,即 $x$ 轴是固定不变的,那么式(5-6-9)是成立的;如果 $x$ 轴不是固定的,即 $x$ 轴的方向是变化的,是时间的函数,那么式(5-6-9)不再成立,即不会出现部分守恒情况。这里"固定"两个字是十分重要的,显然这是动量守恒定律部分守恒情况中一个值得注意的问题。所以部分守恒情况的正确表述应为:如果作用在质点组上的外力作用在某一固定方向上的投影之和恒为零,那么质点组的动量在该方向的投影将保持一常数。

**3. 关于动量守恒条件的判定**

动量守恒定律是由动量定理的特例得到的，因此应有“惯性参考系”的条件。如果将动量定理推广到非惯性系，还需考虑惯性力的冲量。

但是，如果由质点组动量定理的积分形式出发去讨论其条件，容易出现误解。

因为 $\int_{t_0}^{t}\left(\sum \boldsymbol{F}_i\right)\cdot \mathrm{d}t=\boldsymbol{p}-\boldsymbol{p}_0$，为使右方 $\boldsymbol{p}=\boldsymbol{p}_0$，则 $\int_{t_0}^{t}\left(\sum \boldsymbol{F}_i\right)\cdot \mathrm{d}t=0$ 是 $\boldsymbol{p}=\boldsymbol{p}_0$ 的条件。

但是这不是 $\boldsymbol{p}$ 守恒的条件，因为 $\boldsymbol{p}=\boldsymbol{p}_0$ 仅是始、末两个时刻动量相等，并不能保证每一瞬时 $\boldsymbol{p}$ 都不变(即 $\boldsymbol{p}$ 守恒)。

因此，$\int_{t_0}^{t}\left(\sum \boldsymbol{F}_i\right)\cdot \mathrm{d}t=0$，不是动量守恒的条件。故应由动量定理的微分形式讨论：

因为 $\sum_{i=1}^{N}\boldsymbol{F}_i\cdot \mathrm{d}t=\mathrm{d}\boldsymbol{p}$，而动量守恒，即为 $\mathrm{d}\boldsymbol{p}=0$(每瞬时 $\boldsymbol{p}$ 都不变)，则必须 $\sum_{i=1}^{N}\boldsymbol{F}_i\cdot \mathrm{d}t=0$。又由 $\mathrm{d}t\neq 0$，所以 $\sum_{i=1}^{N}\boldsymbol{F}_i=0$ 是动量守恒严格成立的条件。

**4. 动量守恒条件一些补充说明**

(1) 成立条件与系统本身所划定的范围有关。只要适当地扩大范围而划定系统，则总能较好地将外力化为内力，而满足 $\sum \boldsymbol{F}_i=0$ 的条件。

(2) 注意动量是矢量，因此质点组的动量是矢量和不是数量和。所以，$\boldsymbol{p}$ 守恒，从始末两个时刻来看，是 $\sum \boldsymbol{p}_{i0}=\sum \boldsymbol{p}_i$，但不见得满足 $\sum p_{i0}=\sum p_i$，即不能从数值上相等来判断是否守恒。

(3) $\sum \boldsymbol{F}_i=0$ 与“外力作用相互抵消”或“合外力等于零($\boldsymbol{F}_{合}=0$)”不是同一含义。否则条件过窄(只对质点和少数质点组而言)。

“合外力”的概念对质点存在，而对质点组一般并不存在这一概念。“外力的矢量和”即“主矢”的概念不仅对质点组，而且对质点也存在。但仅对质点及少数质点组，这两个概念才有同一意义。

(4) 动量守恒定律近似计算。当内力≫外力时，虽然外力本身看来可能并不算小，但从内、外力对各质点动量的影响相比。外力可以忽略而影响不大，此时，用动量守恒近似计算。这给我们解题提供了方便。

(5) 较常应用的是动量的分量守恒条件，此时总动量不一定守恒。

例如，如果 $\sum F_{ix}=0$，则 $\sum p_{ix}=$恒量。对 $y$ 及 $z$ 方向也可有类似的讨论。

(6) 经典力学中，只讨论实物的动量。实际上的场也有动量。因此，在其他研究领域中，如果过程中既有实物又有场的运动，则动量守恒指“实物动量＋场的动量”之和的守恒，此时只看实物的动量则前后可能并不守恒。

## 5.6.2　机械能守恒定律

**1. 机械能守恒定律的推导**

质点系的动能定理可表示为

$$A_{\mathrm{ext}}+A_{\mathrm{int}}=E_{\mathrm{k1}}-E_{\mathrm{k2}} \tag{5-6-10}$$

质点系中内力分为保守力和非保守力两种,内力做的功包括保守力功和非保守力功

$$A_{int}=A_{int,cons}+A_{int,ircons} \tag{5-6-11}$$

由势能的定义,保守力做的功等于系统内与此保守力相对应的势能的减少

$$A_{int,cons}=E_{p1}-E_{p2} \tag{5-6-12}$$

将式(5-6-11)和式(5-6-12)代入式(5-6-10),可得 $A_{ext}=A_{int,ircons}+(E_{p1}-E_{p2})=E_{k1}-E_{k2}$。改写为

$$A_{ext}+A_{int,ircons}=(E_{k2}+E_{p2})-(E_{k1}+E_{p1}) \tag{5-6-13}$$

质点系的总动能和总势能之和称为质点系的机械能,通常用 $E$ 表示

$$E=E_k+E_p \tag{5-6-14}$$

则式(5-6-13)变为

$$A_{ext}+A_{int,ircons}=E_2-E_1 \tag{5-6-15}$$

式(5-6-15)表明,质点系在运动过程中,外力做的功与系统内非保守力做的功的总和等于它的机械能的增量,这就是质点系的功能原理。当式(5-6-15)中等号左边为0,可以推出机械能守恒定律。此时,左边等于0,即 $A_{ext}+A_{int,ircons}=0$,此就是机械能守恒定律的守恒条件。即

$$E_2-E_1=0 \quad 或 \quad E_2=E_1 \tag{5-6-16}$$

式(5-6-16)表明在质点系的运动过程中,如果满足一定的守恒条件,则质点系的机械能就一定守恒。

**2. 机械能守恒条件的分析**

1)问题的提出

物理量的守恒,通常是指该物理量在整个物理过程中始终保持不变的过程。在研究质点系机械能守恒的过程中,守恒条件是确定其适用范围的关键。在大学物理中,机械能守恒定律的守恒条件有三种典型的提法。

(1)当质点系所受的外力做的功和内力中非保守力做的功之和等于0时,质点系的机械能守恒。即当 $A_{ext}+A_{int,ircons}=0$ 时,$E_k+E_p=C$,$C$ 为恒量。

(2)当质点系所受的外力等于0,内力中非保守力等于0时,质点系的机械能守恒。即当 $\boldsymbol{F}_{ext}=0$,$\boldsymbol{F}_{int,ircons}=0$ 时,$E_k+E_p=C$,$C$ 为恒量。

(3)当质点系所受的外力做的功等于0,内力中非保守力做的功等于0时,质点系的机械能守恒。即当 $A_{ext}=0$,$A_{int,ircons}=0$ 时,$E_k+E_p=C$,$C$ 为恒量。

现对上述提到的各种守恒条件进行讨论。

能量的转化通过做功来实现,功是能量转化的量度。由式(5-6-10)和式(5-6-11)可以看出,外力和内力(非保守力、保守力)做功都会引起质点系动能的变化;由式(5-6-13)可以看出只有外力和非保守力做功时,才能引起系统机械能的改变。特别要注意内保守力做功所起的作用——内保守力做功能够引起质点系动能的改变,不会引起质点系机械能的改变。由此可推知,内保守力做功所引起的系统动能的变化与在此过程中和该保守力相对应的势能的变化刚好相反。对一个质点系,内保守力做功意味着动能和势能的相互转化。

由式(5-6-15)可以看出,只有式(5-6-15)的左边等于0,质点系的机械能才守恒。式(5-6-17)表示机械能守恒条件可能存在的几种情况

$$A_{ext}+A_{int,ircons}=0\begin{cases}A_{ext}=-A_{int,ircons}\\ A_{ext}=0,A_{int,ircons}=0\begin{cases}\boldsymbol{F}_{ext}\neq 0,\boldsymbol{F}_{int,ircons}\neq 0\\ \boldsymbol{F}_{ext}=0,\boldsymbol{F}_{int,ircons}=0\end{cases}\end{cases} \tag{5-6-17}$$

2）$A_{ext}=-A_{int,ircons}$ 情况

假想质点系在所研究的始末状态之间所包含的运动过程中，前阶段外力做正功，大小为 $A_{ext}$，内非保守力不做功，在这个阶段外力做正功，使得质点系的动能增加，势能不变；后阶段外力不做功，而内非保守力做负功，大小也为 $A_{ext}$，即 $A_{ext}=-A_{int,ircons}$，在这个阶段内非保守力做负功使得质点系的动能减少，势能不变。在整个运动过程中，虽然满足了 $A_{ext}+A_{int,ircons}=0$，也只能说始、末状态机械能相等，而不能说这个过程中机械能守恒。守恒是指在运动或变化过程中任意两个状态的机械能都相等，很显然它满足不了这个条件。因此，在这整个过程中，质点系的机械能不守恒，所以 $A_{ext}+A_{int,ircons}=0$ 不是机械能守恒条件的正确提法。

3）$A_{ext}=0$，$A_{int,ircons}=0$ 中 $\boldsymbol{F}_{ext}=0$，$\boldsymbol{F}_{int,ircons}=0$ 情况

一个质点系，在其运动过程中，如果 $\boldsymbol{F}_{ext}=0$，必有 $A_{ext}=\int \boldsymbol{F}_{ext}\cdot \mathrm{d}\boldsymbol{r}=0$；如果 $\boldsymbol{F}_{int,ircons}=0$，必有 $A_{int,ircons}=\int \boldsymbol{F}_{int,ircons}\cdot \mathrm{d}\boldsymbol{r}=0$。

因此，该过程机械能一定守恒。但如果将它作为机械能守恒的条件，就会过于苛刻，无形之中缩小了机械能守恒定律的适用范围。

4）$A_{ext}=0$，$A_{int,ircons}=0$ 中 $\boldsymbol{F}_{ext}\neq 0$，$\boldsymbol{F}_{int,ircons}\neq 0$ 情况

这个条件允许系统在运动过程中受到的外力和内非保守力不为0，但在运动过程中，根据守恒的意义，要求其在任意时间间隔内都不做功，即在任意的时间元 $\mathrm{d}t$ 内做的功 $\mathrm{d}A=0$，也可以写成 $\mathrm{d}A/\mathrm{d}t=0$，而功率 $p=\mathrm{d}A/\mathrm{d}t$，上面的守恒条件相应地变为 $p_{ext}=\frac{\mathrm{d}A_{ext}}{\mathrm{d}t}=0$，$p_{int,ircons}=\frac{\mathrm{d}A_{int,ircons}}{\mathrm{d}t}=0$，这足以说明在质点系的运动过程中，任意两个位置的机械能都相等，即机械能守恒。

综上所述，可以认为机械能守恒定律的守恒条件严格地说应为 $p_{ext}=0$，$p_{int,ircons}=0$。

即当质点系所受外力的功率等于零，内力中非保守力的功率等于零时，质点系的机械能守恒。

### 5.6.3　角动量定理和角动量守恒定律

#### 1. 质点的角动量定理和角动量守恒定律

考虑质量为 $m$ 的质点，在某时刻质点的空间位置矢量为 $\boldsymbol{r}$，用 $\boldsymbol{r}$ 叉乘牛顿第二定律等式两边，则有：$\boldsymbol{r}\times\boldsymbol{F}=\boldsymbol{r}\times\frac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}\times m\boldsymbol{v})-\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\times m\boldsymbol{v}$，因为 $\boldsymbol{v}=\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}$ 和 $\boldsymbol{v}\times\boldsymbol{v}=0$，所以 $\boldsymbol{r}\times\boldsymbol{F}=\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}\times m\boldsymbol{v})$。式中 $\boldsymbol{r}\times\boldsymbol{F}$ 为作用在质点上的合力对参考点（位置矢量的原点）的力矩，用 $\boldsymbol{M}$ 表示，即 $\boldsymbol{M}=\boldsymbol{r}\times\boldsymbol{F}$；$\boldsymbol{r}\times m\boldsymbol{v}$ 定义为对同一参考点的角动量，用 $\boldsymbol{L}$ 表示，即 $\boldsymbol{L}=\boldsymbol{r}\times m\boldsymbol{v}$。则有

$$\boldsymbol{M}=\frac{\mathrm{d}\boldsymbol{L}}{\mathrm{d}t} \tag{5-6-18}$$

相对于惯性系中的一个固定参考点，如果作用在质点上的合力矩为零，即 $\boldsymbol{M}=0$，质点对该固定点的角动量守恒，即

$$\boldsymbol{L}=\boldsymbol{r}\times m\boldsymbol{v}=\text{常量} \tag{5-6-19}$$

**2. 质点系的角动量定理和角动量守恒定律**

质点系的角动量定理可通过对各质点所遵从的角动量定理求和得到。考虑由 $N$ 个质点组成的质点系,对其中第 $i$ 个质点,角动量定理要表示为

$$\boldsymbol{r}_i \times \boldsymbol{F}_i = \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}_i \times m_i \boldsymbol{v}_i) \tag{5-6-20}$$

1) 质点系的角动量定理

对于 $N$ 个质点有:$\sum \boldsymbol{r}_i \times \boldsymbol{F}_i = \sum \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}_i \times m_i \boldsymbol{v}_i)$。由于内力力矩相互抵消,$\sum \boldsymbol{r}_i \times \boldsymbol{F}_i$ 中只有外力矩有贡献,故有 $\sum \boldsymbol{r}_i \times \boldsymbol{F}_{i外} = \sum \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}_i \times m_i \boldsymbol{v}_i)$。令 $\boldsymbol{M}_{外} = \sum \boldsymbol{r}_i \times \boldsymbol{F}_{i外}$,$\boldsymbol{L} = \sum (\boldsymbol{r}_i \times m_i \boldsymbol{v}_i)$,则得质点系的角动量定理:

$$\boldsymbol{M}_{i外} = \frac{\mathrm{d}\boldsymbol{L}}{\mathrm{d}t} \tag{5-6-21}$$

2) 质点系角动量守恒条件的判定

当外力矩为零时,即 $\boldsymbol{M}_{外}=0$,得质点系的角动量守恒定律:

$$\boldsymbol{L} = \sum (\boldsymbol{r}_i \times m_i \boldsymbol{v}_i) = 常量 \tag{5-6-22}$$

由质点组角动量定理的微分形式,因 $\boldsymbol{M}_{i外} = \frac{\mathrm{d}\boldsymbol{L}}{\mathrm{d}t}$,所以当 $\boldsymbol{M}_{i外}=0$ 时(外力矩的矢量和为零,并不要求每个外力矩必为 0),任一时刻 $\boldsymbol{L}$=恒矢量。

3) 质点系角动量守恒条件补充说明

(1) 成立条件与系统本身所划定的范围有关。适当扩大范围,一般总能满足守恒条件。

(2) 仍须注意角动量守恒也是矢量守恒。

(3) "外力矩的矢量和"不是"外力矢量和的力矩"。对质点组后者也根本没有物理意义。

(4) 用角动量守恒定律近似计算的条件较多用于处理有转动物体参与的强烈碰撞问题。

(5) 分量守恒条件:当 $M_{iz}=0$ 时,$L_z$=恒量。(对 $x$ 及 $y$ 方向的分量可做类似讨论。)

(6) 守恒条件还与参考点的选择有关。因为 $\boldsymbol{M}_{i外}$ 与参考点有关。此与动量守恒不同。

## 5.7 受迫振动中的位移共振与速度共振

物体在平衡位置附近往复的运动叫作振动,或机械振动。波是振动的传播,机械振动的传播即机械波。掌握振动的规律是研究波动的基础。我们通常研究的是简谐振动。从动力学的角度,简谐振动产生是由于系统只受到与形变大小成正比例、方向与形变方向相反的力的作用;从运动学的角度,简谐振动是振幅不变的等幅振荡;从能量的角度,简谐振动是振动能量没有损耗的振动。但是,实际中的振动系统,必然有能量损耗,所以,我们必须对振动系统补充能量,即形成受迫振动。

### 5.7.1 受迫振动

振动系统在连续的周期性外力(亦称驱动力)作用下进行的振动称为受迫振动。

**1. 受迫振动的动力学方程和运动学方程**

受迫振动的质点受三种力：弹性力$-kx$，阻尼力$-\gamma\dfrac{dx}{dt}$，周期性外力$F(t)=F_0\cos\omega t$（初相位为零）。根据牛顿第二定律，受迫振动的动力学方程式为

$$m\frac{d^2x}{dt^2}=-kx-\gamma\frac{dx}{dt}+F_0\cos\omega t \tag{5-7-1}$$

为方便起见，令$\omega_0^2=\dfrac{k}{m}$，$2\beta=\dfrac{\gamma}{m}$，$f_0=\dfrac{F_0}{m}$，得

$$\frac{d^2x}{dt^2}+2\beta\frac{dx}{dt}+\omega_0^2x=f_0\cos\omega t \tag{5-7-1$'$}$$

这是受迫振动动力学方程常见的形式。根据微分方程的理论，方程(5-7-1)′的解为

$$x=A_0e^{-\beta t}\cos(\omega' t+\alpha)+A\cos(\omega t+\varphi) \tag{5-7-2}$$

式(5-7-2)是受迫振动运动学方程，其中的$A$和$\alpha$是由初始条件决定的积分常数，$\omega'=\sqrt{\omega_0^2-\beta^2}$。此解为两项之和，表明质点运动包含两个分运动。第一项为阻尼振动，随时间的推移而趋于消失，它反映受迫振动的暂态行为，与驱动力无关。第二项表示与驱动力频率相同且振幅为$A_0$的周期振动。振动开始时，受迫振动的振幅较小，经过一定时间后，阻尼振动既可忽略不计，质点进行由式(5-7-2)第二项所决定的与驱动力同频率的振动，称为受迫振动的稳定振动状态。

**2. 稳定受迫振动的特征量**

因为$t\to\infty$时，$e^{-\beta t}\to 0$，所以，稳定受迫振动的振动方程为

$$x=A\cos(\omega t+\varphi) \tag{5-7-3}$$

则，稳定受迫振动的圆频率与驱动力的圆频率相同，振幅和初位相分别为

$$A=\frac{f_0}{\sqrt{(\omega_0^2-\omega^2)^2+4\beta^2\omega^2}} \tag{5-7-4}$$

$$\tan\varphi=\frac{-2\beta\omega}{\omega_0^2-\omega^2} \tag{5-7-5}$$

显然，当$\omega=0$时，$\varphi=0$；当$\omega\leqslant\omega_0$时，$0>\varphi\geqslant-\dfrac{\pi}{2}$；当$\omega\geqslant\omega_0$时，$-\dfrac{\pi}{2}\geqslant\varphi\geqslant-\pi$。由此可见，稳定的受迫振动的位移变化总是滞后于驱动力一个$\varphi$角。

## 5.7.2　位移共振与速度共振的对比

共振是力学、电磁学、光学、原子物理学等学科中最常遇到的一种现象，一些教科书中把共振定义为：受迫振动的振幅达到最大值的现象。其实，这样的共振，表明的只是位移达到了最大值的一种共振，应该准确地称之为位移共振，而像振动速度达到最大值的共振，则应称之为速度共振。

**1. 共振条件的比较**

如果要使受迫振动的位移达到最大值，可以令$\dfrac{dA}{d\omega}=0$，即$\dfrac{d}{d\omega}\left[\dfrac{f_0}{\sqrt{(\omega_0^2-\omega^2)^2+4\beta^2\omega^2}}\right]=\dfrac{2\omega f_0(\omega_0^2-\omega^2-2\beta^2)}{[(\omega_0^2-\omega)^2+4\beta^2\omega^2]^{3/2}}=0$，从而得到

$$\omega_r=\sqrt{\omega_0-2\beta^2} \tag{5-7-6}$$

式(5-7-6)就是发生位移共振时,驱动力圆频率 $\omega$ 要满足的条件。同理,由受迫振动的速度振幅

$$v=A\omega=\frac{\omega f_0}{\sqrt{(\omega_0^2-\omega^2)^2+4\beta^2\omega^2}} \tag{5-7-7}$$

令$\dfrac{\mathrm{d}V}{\mathrm{d}\omega}=0$,可得

$$\omega_r=\omega_0 \tag{5-7-8}$$

式(5-7-8)就是发生速度共振时,驱动力圆频率 $\omega$ 应满足的条件。

**2. 共振相位的比较**

在位移共振中,将共振条件 $\omega_r=\sqrt{\omega_0-2\beta^2}$ 代入初位相式(5-7-5)可得 $\varphi=\arctan\dfrac{-\sqrt{\omega_0^2-2\beta^2}}{\beta}$,且由于 $\omega<\omega_0$,所以 $0>\varphi>-\dfrac{\pi}{2}$,即 $\varphi$ 位移变化滞后于驱动力变化的位相。

位移共振时速度与驱动力并不同相。因此在位移的一个周期内,有时驱动力与速度同方向,驱动力对振动系统做正功;有时驱动力与速度反方向,驱动力对振动系统做负功,其差值用于振动系统克服阻力做功而转化为热能。

在速度共振中,将共振条件 $\omega_r=\omega_0$ 代入式(5-7-5)可得 $\varphi=-\dfrac{\pi}{2}$,表明位移变化总是滞后于驱动力$\dfrac{\pi}{2}$位相,比位移共振中的位移变化更为滞后。将 $\omega_r=\omega_0$ 代入式(5-7-7)可得

$$v=\frac{f_0}{2\beta}=\frac{f_0}{2\beta}\cos\omega t \tag{5-7-9}$$

式(5-7-9)指出,速度共振时速度与驱动力同相。因此在速度共振的一个周期内,驱动力总是与速度同方向,驱动力总是对振动系统做正功。驱动力做的功完全用于振动系统克服阻力做功而转化为热能。

**3. 共振振幅的比较**

在位移共振中,将位移速度共振条件代入式(5-7-4),得其振幅为

$$A=\frac{f_0}{2\beta\sqrt{\omega_0^2-\beta^2}} \tag{5-7-10}$$

同样,将速度共振条件代入式(5-7-4),便可得到速度共振时的振幅为

$$A=\frac{f_0}{\sqrt{4\omega^2\beta^2}}=\frac{f_0}{2\omega\beta} \tag{5-7-11}$$

显然,位移共振和速度共振中的振幅是不相同的,后者偏小。

在位移共振中,有 $F(t)=F_0\cos\omega t$,$x=A\cos(\omega t+\varphi)$ 和 $v=-\omega A\sin(\omega t+\varphi)=\omega A\cos\left(\omega t+\varphi+\dfrac{\pi}{2}\right)$。由于 $0>\varphi>-\dfrac{\pi}{2}$,所以,$0<\left(\varphi+\dfrac{\pi}{2}\right)<\dfrac{\pi}{2}$。这表明在位移共振中速度的变化总是超前驱动力的变化。驱动力的瞬时功率为

$$P_{强}=F(t)v=F_0\cos\omega t\cdot\omega A\cos\left(\omega t+\varphi+\frac{\pi}{2}\right)$$
$$=F_0\omega A\cos\omega t\left[\cos\omega t\cos\left(\varphi+\frac{\pi}{2}\right)-\sin\omega t\sin\left(\varphi+\frac{\pi}{2}\right)\right]$$
$$=F_0\omega A\cos^2\omega t\cos\left(\varphi+\frac{\pi}{2}\right)-F_0\omega A\cos\omega t\sin\omega t\sin\left(\varphi+\frac{\pi}{2}\right),$$

因为，$\frac{1}{T}\int_0^T\cos^2\omega t\,\mathrm{d}t=\frac{1}{2}$，$\frac{1}{T}\int_0^T\cos\omega t\sin\omega t\,\mathrm{d}t=0$，所以平均功率

$$\bar{P}_{强}=\frac{1}{T}\int_0^T P_{强}\,\mathrm{d}t=\frac{1}{2}F_0\omega A\cos\left(\varphi+\frac{\pi}{2}\right)=-\frac{1}{2}F_0\omega A\sin\varphi$$
$$=-\frac{1}{2}F_0\omega\cdot\frac{f_0}{\sqrt{(\omega_0^2-\omega^2)^2+4\beta^2\omega^2}}\cdot\sin\left(\arctan\frac{-2\beta\omega}{\omega_0^2-\beta^2}\right)$$
$$=-\frac{1}{2}f_0m\omega\cdot\frac{f_0}{\sqrt{(\omega_0^2-\omega^2)^2+4\beta^2\omega^2}}\cdot\frac{\frac{-2\beta\omega}{\omega_0^2-\omega^2}}{\sqrt{1+\left(\frac{-2\beta\omega}{\omega_0^2-\omega^2}\right)^2}}=\frac{f_0^2\omega^2m\beta}{(\omega_0^2-\omega^2)^2+4\beta^2\omega^2}$$

在一个周期内驱动力做的功：$W_{强}=P_{强}\ T=\frac{2\pi f_0^2\omega m\beta}{(\omega_0^2-\omega^2)^2+4\beta^2\omega^2}$。相反，振动系统克服阻尼力做功，其平均功率

$$\bar{P}_{阻}=\frac{1}{T}\int_0^T fv\,\mathrm{d}t=\frac{1}{T}\int_0^T-\gamma v^2\,\mathrm{d}t=\frac{1}{T}\int_0^T-2m\beta\cdot\omega^2A^2\cos^2\left(\omega t+\varphi+\frac{\pi}{2}\right)\mathrm{d}t$$
$$=-m\beta\omega^2A^2=-\frac{f_0^2\omega^2m\beta}{(\omega_0^2-\omega^2)^2+4\beta^2\omega^2}$$

可见，在位移共振中，驱动力对系统做的功与系统克服阻尼力做的功相等，也即系统吸收的能量与消耗的能量相等，从而保持了稳定的振动特点。

在速度共振中，$F(t)=F_0\cos\omega t$ 和 $x=A\cos(\omega t+\varphi)$。由于 $\varphi=-\frac{\pi}{2}$，所以

$$v=-\omega A\sin(\omega t+\varphi)=\omega A\cos\left(\omega t+\varphi+\frac{\pi}{2}\right)=\omega A\cos\omega t \qquad (5\text{-}7\text{-}12)$$

式(5-7-12)表明：在速度共振中，速度的变化与驱动力的变化位相相同，驱动力一直做正功，其平均功率

$$\bar{P}_{强}=\frac{1}{T}\int_0^T Fv\,\mathrm{d}t=\frac{1}{T}\int_0^T F_0\cos\omega t\cdot\omega A\cos\omega t\,\mathrm{d}t$$
$$=F_0\omega A\cdot\frac{1}{T}\int_0^T\cos^2\omega t\,\mathrm{d}t=\frac{1}{2}F_0\omega A=\bar{P}_{max}$$
$$=\frac{1}{2}f_0m\omega\cdot\frac{f_0}{2\beta\omega}=\frac{f_0^2m}{4\beta}$$

一个周期内做的功，$W=PT=\frac{f_0^2m}{4\beta}\cdot\frac{2\pi}{\omega_0}=\frac{\pi f_0^2m}{2\beta\omega_0}$；相反，系统克服阻力做功，其平均功率

$$\bar{P}_{阻}=\frac{1}{T}\int_0^T fv\,\mathrm{d}t=\frac{1}{T}\int_0^T -\gamma v^2\,\mathrm{d}t=\frac{1}{T}\int_0^T -\gamma\omega^2A^2\cos^2\omega t\,\mathrm{d}t$$

$$=-\frac{1}{2}\gamma\omega^2A^2=-\frac{1}{2}2\beta m\omega^2\frac{f_0^2}{4\beta^2\omega^2}=-\frac{f_0^2m}{4\beta}=-\bar{P}_{强}$$

可见，在速度共振中，驱动力对系统做的功最多，提供的能量最大，但由于这时振动速度最大，阻力也最大，阻力做的功也最多，且与驱动力做的功相等，所以速度共振系统仍是一个稳定的振动系统。

### 5.7.3 总结

由以上分析可见，速度共振是区别于通常教科书中所定义的位移共振概念的另一类共振，它是振动速度达到最大值的一种受迫振动，其共振条件是驱动力的圆频率 $\omega$ 与振动系统固有圆频率 $\omega_0$ 相等，即 $\omega_r=\omega_0$，而不像位移共振的条件是 $\omega_r=\sqrt{\omega_0-2\beta^2}$；其振幅 $A=\dfrac{f_0}{2\beta\sqrt{\omega_0^2-\beta^2}}$比位移共振振幅 $A=\dfrac{f_0}{\sqrt{4\omega^2\beta^2}}=\dfrac{f_0}{2\omega\beta}$小；其速度变化的位相总是与驱动力变化的位相相同，而不像位移变化的位相总是滞后于驱动力$\dfrac{\pi}{2}$，因而速度共振系统吸收驱动力的能量最多，$W=\dfrac{\pi f_0^2m}{2\beta\omega_0}$，比位移共振中的能量 $W=\dfrac{2\pi f_0^2\omega m\beta}{(\omega_0^2-\omega^2)^2+4\beta^2\omega^2}$大，但由于在速度共振中，振动速度大，克服阻力消耗的能量也最大，所以对于速度共振系统，仍是一个稳定的共振系统，其物理意义明确，振动特征明显，为此我们应将共振概念扩展为："凡在一定条件下，某一物理量达到最激烈最大值的受迫振动，都应称作共振。"位移达到最大值的共振就叫位移共振，像上面讨论的速度达到最大值的共振，就称作速度共振。

## 5.8 为什么弹性介质波的动能与势能是同步的

机械波动是振动在弹性介质中的传播。波的传播过程既是振动的传播过程，又是能量的传递过程。由于波的传播，介质中的质点发生振动，因而具有动能；同时由于介质元都发生压缩或伸张形变(纵波)或切向形变(横波)，因而具有形变势能。动能和势能的总和就是该介质元的总机械能。因此波传到哪里，哪里就有机械能。这些机械能来自于波源。在介质元不迁移的情况下传递能量是介质波动的基本性质。

### 5.8.1 介质中质点振动能量的计算问题

机械波动就是机械振动的传播过程，也是机械能量的传递过程。波的能量应该就是介质中质点振动的能量。这里只讨论一维简谐波。介质中质点是作与振源(波源)同频率的简谐振动，如果介质无吸收，各质点的振幅也与振源相同。介质中各质点是简谐振动，振动能量似乎可用简谐振动能量公式来计算。简谐振动的总机械能为

$$E=E_k+E_p=\frac{1}{2}kA^2=\frac{1}{2}m\omega^2A^2 \tag{5-8-1}$$

计算结果显示各质点的机械能守恒。照这样说来，波动岂不是根本无法传递能量了吗？

那么，问题出在哪里呢？原来这里对于介质元的总机械能不可套用公式(5-8-1)。对于单独的简谐振子，弹性形变的弹性势能取决于振子本身偏离平衡位置的位移，其机械能即用

公式(5-8-1)计算。至于说到介质，如果一群相邻的质点(严格地说，应当是“形变的介质元”)偏离各自平衡位置的位移相同，换句话说，这一部分介质作整体的移动，那么这部分介质并无形变，从而这一群质点并无弹性势能。必须各质点位移不同，这部分介质才有形变，这群质点才有弹性势能。因此，在介质中，弹性势能并非取决于各质点偏离各自平衡位置的位移，而是取决于相邻质点之间的相对位移，即“形变的介质元”。这使得公式(5-8-1)不再适用。这也是机械波波动动能与势能同步的根本原因。

## 5.8.2　机械波的能量

下面以简谐纵波在一棒中沿棒长方向传播为例，推导出波的能量公式。

如图 5-8-1 所示，取 $x$ 轴沿棒长方向，设波动方程为 $y=A\cos\left(\omega t-\frac{x}{\mu}\right)$。在波动过程中，棒中每一小段将不断地压缩和拉伸。

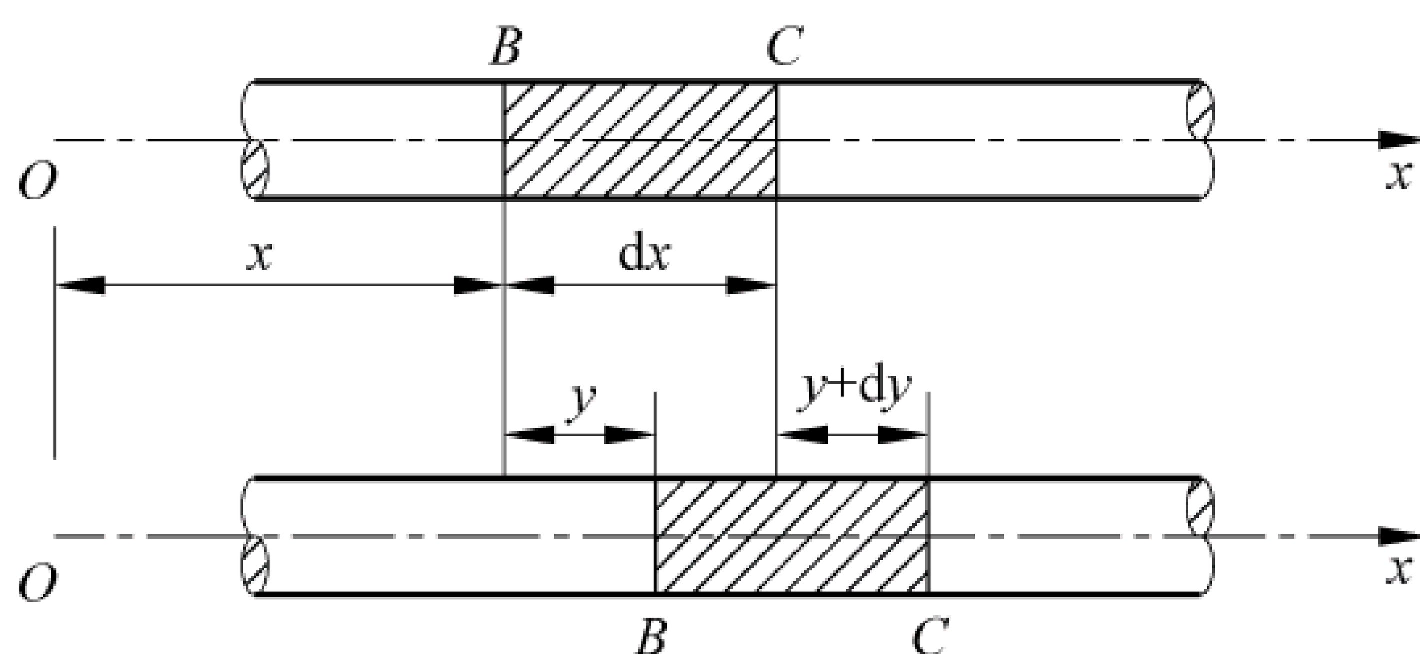



图 5-8-1　棒中波动的分析

在棒上任取一体积元(注意：是体积元，而不是质点)$BC$，体积 $\mathrm{d}V$，棒在平衡位置时，$B$、$C$ 坐标分别为 $x$ 和 $x+\mathrm{d}x$，即 $BC$ 长为 $\mathrm{d}x$。设棒的横截面积为 $S$，质量密度为 $\rho$，体积元能量为 $\mathrm{d}W=\mathrm{d}W_{\mathrm{k}}+\mathrm{d}W_{\mathrm{p}}$。

**1. 动能与势能**

体积元动能为

$$\mathrm{d}W_{\mathrm{k}}=\frac{1}{2}\mathrm{d}mV^2=\frac{1}{2}\rho\mathrm{d}V\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2=\frac{1}{2}\rho\mathrm{d}V\cdot\omega^2A^2\sin^2\omega\left(t-\frac{x}{\mu}\right)\tag{5-8-2}$$

势能由以下推导求得：设 $t$ 时刻，$A$、$B$ 端位移分别为 $y$、$y+\mathrm{d}y$，所以，体积元伸长量为 $\mathrm{d}y$。设在体积元端面上由于形变产生的弹性恢复力大小为 $f$，可知，协强为 $f/S$，协变为 $\mathrm{d}y/\mathrm{d}x$，根据我们熟知的胡克定律，按照杨氏弹性模量定义有

$$\frac{f}{S}=Y\frac{\mathrm{d}y}{\mathrm{d}x}\tag{5-8-3}$$

式中 $Y$ 为杨氏弹性模量。则有

$$f=SY\frac{\mathrm{d}y}{\mathrm{d}x}\tag{5-8-4}$$

按照胡克定律，在弹性限度内弹性恢复力值为

$$f=k\mathrm{d}y\tag{5-8-5}$$

其中 $k=\frac{YS}{\mathrm{d}x}$。由式(5-8-4)、式(5-8-5)和弹性势能的定义，有

$$\mathrm{d}W_{\mathrm{p}}=\frac{1}{2}k(\mathrm{d}y)^2=\frac{1}{2}\cdot YS/\mathrm{d}x\cdot(\mathrm{d}y)^2=\frac{1}{2}YS\mathrm{d}x\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2=\frac{1}{2}Y\mathrm{d}V\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\tag{5-8-6}$$

因为$\mu=\sqrt{\frac{Y}{\rho}}$，所以，$Y=\rho\mu^2$；又因为$y=y(x,t)$，$\frac{\mathrm{d}y}{\mathrm{d}x}$应写成$\frac{\partial y}{\partial x}$，可有$\frac{\partial y}{\partial x}=\frac{\omega}{\mu}A\sin\omega\left(t-\frac{x}{\mu}\right)$，得体积元势能为

$$\mathrm{d}W_{\mathrm{p}}=\frac{1}{2}\rho\mu^2\cdot\mathrm{d}V\left[\frac{\omega^2}{\mu^2}A^2\sin^2\omega\left(t-\frac{x}{\mu}\right)\right]=\frac{1}{2}\rho\mathrm{d}V\omega^2A^2\sin^2\omega\left(t-\frac{x}{\mu}\right)\tag{5-8-7}$$

可得总能量为

$$\mathrm{d}W=\mathrm{d}W_{\mathrm{k}}+\mathrm{d}W_{\mathrm{p}}=\rho\mathrm{d}V\omega^2A^2\sin^2\omega\left(t-\frac{x}{\mu}\right)\tag{5-8-8}$$

**2. 讨论**

(1) 任一时刻体积元动能与其势能总是相等：

$$\mathrm{d}W_{\mathrm{k}}=\mathrm{d}W_{\mathrm{p}}=\frac{1}{2}\rho\mathrm{d}V\omega^2A^2\sin^2\omega\left(t-\frac{x}{\mu}\right)\tag{5-8-9}$$

(2) 波动中体积元的能量与单一谐振动系统的能量有着显著的不同。

在单一谐振动的系统中，动能和势能相互转化，动能最大时，势能最小，势能最大时，动能最小，系统机械能守恒。但是波动的任一时刻任一体积元的动能与势能总是随时间变化的，变化是同步的，值也相等，这说明体积元总能量不能为常数，即能量不守恒(体积元)。

(3) 波动中体积元能量不守恒的原因

每个体积元都不是独立地作谐振动，它与相邻的体积元间有着相互作用，因而相邻体积元间有能量传递。沿着波传播方向，某体积元从前面介质获得能量，又把能量传递给后面介质，这样，通过体积元不断地吸收和不断传递能量。所以波动是能量传递的一种形式。

### 5.8.3 波的能量密度

为了描述波动过程中能量的传播，必须引入能量密度的概念。波动的能量密度$w$即单位体积内波动能量：

$$w=\frac{\mathrm{d}W}{\mathrm{d}V}=\rho\omega^2A^2\sin^2\omega\left(t-\frac{x}{\mu}\right)\tag{5-8-10}$$

可知，$w$是$t$的函数。能量密度随时间周期性变化，其周期为波动周期的一半。

平均能量密度$\bar{w}$：

$$\bar{w}=\frac{1}{T}\int_0^T w\,\mathrm{d}t=\frac{1}{T}\int_0^T\rho\omega^2A^2\sin^2\omega\left(t-\frac{x}{\mu}\right)\mathrm{d}t=\rho\omega^2A^2\ \frac{1}{T}\int_0^T\ \frac{1}{2}\left[1-\cos2\omega\left(t-\frac{x}{\mu}\right)\right]\mathrm{d}t$$

$$=\rho\omega^2A^2\ \frac{1}{T}\left[\frac{1}{2}T-\frac{1}{2}\int_0^T\cos2\omega\left(t-\frac{x}{\mu}\right)\mathrm{d}t\right]=\frac{1}{2}\rho\omega^2A^2$$

### 5.8.4 波动中的质点与弹簧振子简谐振动的对比

从运动学角度考虑，波动中的体积元的振动确实与以弹簧振子(质点)为代表的简谐振动相同。但从动力学角度看，两者却有很大的不同。弹簧振子系统中，物块$m$受到系统内力——弹簧的弹性力作用，弹性力是保守力，它做功使弹性势能减少，物块动能增加；反之，物块克服弹性力做功，弹性势能增加，物块动能减少，动能与势能互相转换。同时，弹簧振子系统是一个孤立系统(质点)，并未与外界交换能量(弹簧固定端虽与外界接触，但接触点不运动，不做功)。因此系统机械能守恒。

介质中的质点体积元则不同，它的形变取决于相邻质点的相对位移，弹性力也是体积元

相邻质点之间的相互作用力。就每个质点而言,这弹性力是外力,从而每个体积元的质点并非孤立系统。在平衡位置,质点的动能达极大值,而从波形曲线上看,平衡位置斜率也极大(图 5-8-2 中的 $B$ 点),处于平衡位置的体积元 $dV$ 的形变也极大,势能因此也极大。在最大偏离处(图 5-8-2 中的 $A$ 点),体积元动能为零,而波形曲线的斜率也为零,$dV$ 的形变也为零,因此势能也为零。这样,各质点的机械能并不守恒。其实,这是理当如此的,既然体积元各质点并非孤立系统,体积元相邻质点相互做功,从而交换能量,各质点的能量当然就不可能守恒,能量在介质的质点之间传播。

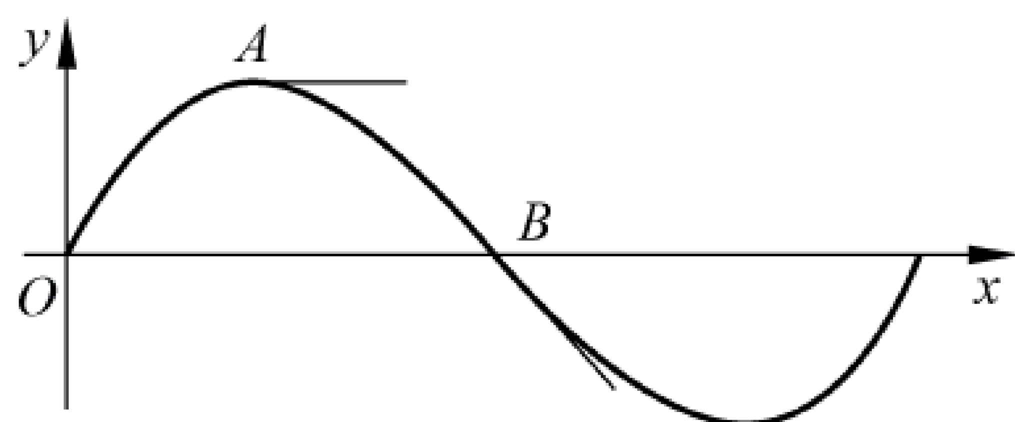


图 5-8-2　波形

### 5.8.5　机械简谐波的波形及其传播示意图

**1. 介质元的线变产生纵波**

弹性介质元的线变产生纵波,如图 5-8-3 所示。纵波内介质质元的线变——波的平衡位置,形变最大、速度最大,在波峰位置,形变最小,速度为零。

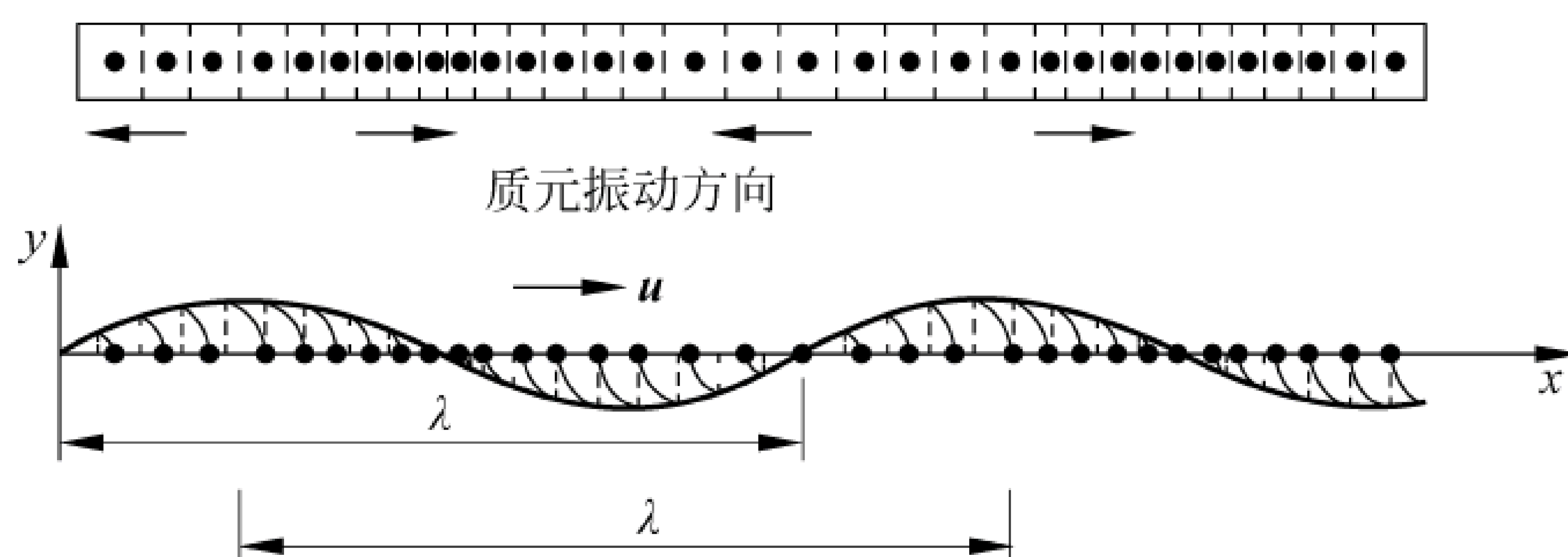


图 5-8-3　纵波示意图

**2. 介质元的切变产生横波**

弹性介质元切变产生横波,如图 5-8-4 所示。横波内介质质元的切变——波的平衡位置,形变最大,速度也最大,在波峰位置,形变最小,速度为零。

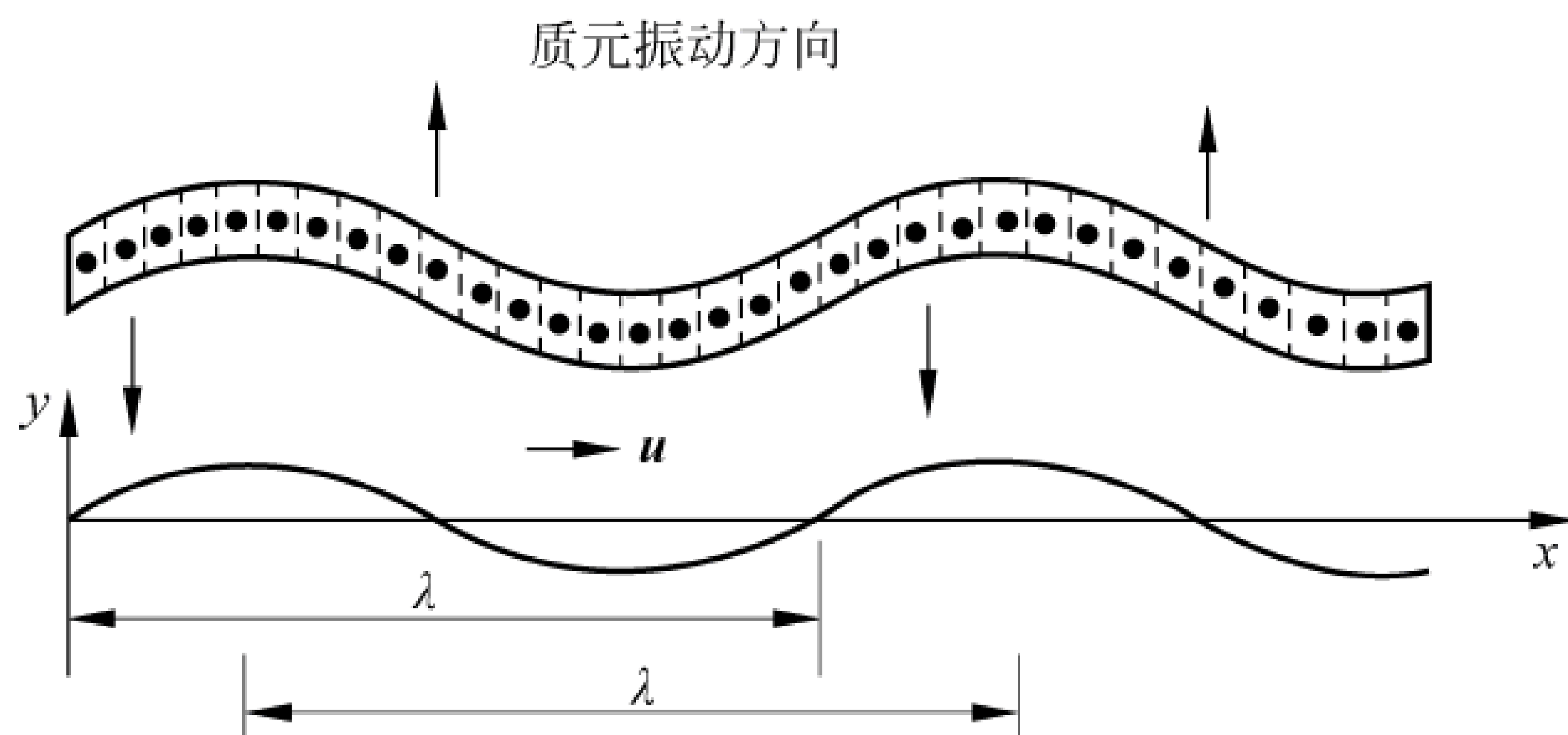


图 5-8-4　横波示意图

## 5.9　经典力学的成就和局限性

本节首先介绍经典力学的伟大成就,简单介绍揭开现代物理学序幕的迈克耳孙-莫雷实验和黑体辐射实验及其意义,指出经典力学的局限性及经典力学的适用范围。

### 5.9.1　经典力学的伟大成就

15 世纪以后,物理学进入了一个迅速发展的阶段。16 世纪哥白尼(Nicolaus Copernicus,

1473—1543)建立了日心说。17世纪伽利略发现了惯性定律。牛顿在伽利略、笛卡儿、开普勒、惠更斯等人研究的基础上，总结出了牛顿运动定律和万有引力定律。17、18世纪科学家完善了经典力学体系。

经典力学的建立标志着近代自然科学的诞生。经典力学把天上和地面物体的运动统一起来，实现了人类对自然界认识的第一次理论大综合。在研究方法上，人们把经典力学中行之有效的实验和数学相结合的方法推广到物理学的各个分支科学上，相继建立了热力学、声学、光学和电动力学等，从而形成了完整的经典物理学体系。

经典力学可以解决低速、宏观条件下几乎所有的力学问题，小到质点的运动规律，大到天体运行规律、宇宙飞船运动的掌控，其伟大成就是人类文明的卓越标志。相对论的建立是与经典力学的局限性紧密联系的。经典力学的局限性仅仅在于其适用条件的特定性，其理论本身没有任何的不自洽和缺陷，经典力学的伟大成就是不容置疑的。

### 5.9.2 经典力学的局限性

由于经典力学在理论和实践上的巨大成功，在18世纪出现了力图把力学规律和它对自然现象的描述方法推广到一切自然研究活动中去的机械自然观。这种观念，对近代自然科学的发展产生了深刻而广泛的影响，几乎所有的物理学家都把经典力学看作是所有其他自然科学的基础。其实，经典力学和其他任何理论一样，有其自身的局限性和适用范围。

1900年的4月27日，在英国伦敦皇家研究所举行的科学报告会上，一位德高望重的老科学家开尔文作了一个演讲，题目是《在热和光动力理论上空的19世纪乌云》，文中指出“动力学理论断言：热和光都是运动的方式，但现在这一理论的优美性和明晰性却被两朵乌云遮蔽，显得黯然失色了……”。

这两朵乌云分别是迈克耳孙-莫雷实验和黑体辐射实验。

**1. 迈克耳孙-莫雷实验**

经典力学的研究表明，所有的机械波都必须通过介质才能传播。麦克斯韦的电磁场理论揭示了光是一种电磁波，光既然是波，当时的人们自然要想：传播光波的介质是什么？当时的物理学家假定这种介质叫作“以太”，整个宇宙空间都充满了“以太”，认为“以太”是非常稀薄、密度极小的、完全透明的、难以感觉到的介质。

物理规律需要用一定参考系表述出来。在经典力学中，根据实践经验引入了惯性参考系。我们知道力学的基本运动定律对所有惯性系成立。按照旧时空概念，如果物质运动速度相对于某一参考系为$c$，则变换到另一参考系时，其速度就不可能沿各个方向都为$c$。从旧概念出发，电磁波只能对一个特定参考系的传播速度为$c$，因而麦氏方程组也就只能对该特殊参考系成立。如果确是这样，则经典力学中一切惯性参考系等价原理的相对性原理在电磁现象中就不再成立。因而由电磁现象可以确定一个特殊参考系，这样便可以把相对于该特殊参考系的运动称为绝对运动。

以往，当我们研究物体的运动以地面作参照物时，地球其实也是在运动的，当我们以太阳作参照物时，太阳也是在运动的，地球和太阳都不是精确的惯性参考系。如果用充满宇宙的“以太”作惯性参考系研究物体的运动，就不会有这种尴尬了，所以相对于“以太”的运动就可以称为“绝对运动”。迈克耳孙实验就是期望找到在绝对静止的参考系中发生的物理现象来证明以太的存在并得到绝对运动的。

迈克耳孙实验根据光干涉原理(详细见第11章)，两束相干光到达观察者眼中，将会出现明暗相间的干涉条纹，若两束光同时到达时中央亮纹位置居中，若两束光到达时间有先

后，就会发生亮纹的旁移；迈克耳孙设想以太存在且完全静止，那么地球相对于以太作绝对运动，对于安装在地面上的仪器而言，相当于以太以地球运动的速率吹过地面，若将仪器转过 90°，两光束相对于以太风的速度会改变，干涉条纹的中央亮纹位置会明显旁移，观察者使用该装置应当容易观察到在绝对参考系下发生的这种干涉条纹移动的现象。但是，迈克耳孙在各种条件下一次又一次地精心实验，却始终未能观察到干涉条纹有丝毫的移动，说明光沿不同方向的传播速度无差异。在惊愕之余，事实不容置辩地指出，以太是不存在的，绝对静止的参考系也是不存在的，真空中的光速对任何参考系都是一个恒量。

爱因斯坦在仔细分析了对迈克耳孙-莫雷实验的各种解说并认真思考后，大胆地认为应该彻底抛弃“以太”学说，建立新的理论——相对论。相对论时空观与经典力学时空观的区别详见表 5-9-1。

**表 5-9-1　相对论时空观与经典力学时空观的区别**

| 项　　目 | 经典物理的时空观 | 相对论时空观 |
|---|---|---|
| 真空中光速 | 在不同惯性参考系中测得光速不同 | 在任何惯性参考系中测得的光速都相同 |
| 介质 | 宇宙中传播光的介质叫“以太” | 宇宙中不存在“以太” |
| 时间 | 时间是独立的，与参考系无关 | 同一事件，不同的参考系中测得时间不同 |
| 长度 | 同一物体的长度与参考系无关 | 同一物体，不同的参考系中测得长度不同 |
| 速度 | 可以任意大 | 真空中的光速是速度的极限 |

**2. 黑体辐射实验（1900 年前后）：研究热辐射的能量分布与温度的关系**

在同样的温度下，不同物体的发光亮度和颜色（波长）不同。颜色深的物体吸收辐射的本领比较强，比如煤炭对电磁波的吸收率可达到 80%左右。所谓“黑体”是指能够全部吸收外来的辐射而无任何反射和透射，吸收率是 100%的理想物体。真正的黑体并不存在，但是，一个表面开有一个小孔的空腔，则可以看作是一个近似的黑体。因为通过小孔进入空腔的辐射，在腔里经过多次反射和吸收以后，不会再从小孔透出。

19 世纪末，卢梅尔（Lummer，1860—1925）等人的著名实验——黑体辐射实验，发现黑体辐射的能量不是连续的，它按波长的分布仅与黑体的温度有关。从经典物理学的角度看来，这个实验的结果是不可思议的。

怎样解释黑体辐射实验的结果呢？当时，人们都从经典物理学出发寻找实验的规律。前提和出发点不正确，最后都导致了失败的结果。例如，德国物理学家维恩建立起黑体辐射能量按波长分布的公式，但这个公式只在波长比较短、温度比较低的时候才和实验事实符合。英国物理学家瑞利和物理学家、天文学家金斯认为能量是一种连续变化的物理量，建立起在波长比较长、温度比较高的时候和实验事实比较符合的黑体辐射公式。但是，从瑞利-金斯公式推出，在短波区（紫外光区）随着波长的变短，辐射强度可以无止境地增加，这是根本不可能的。所以这个失败被埃伦菲斯特称为“紫外灾难”。它的失败无可怀疑地表明经典物理学理论在黑体辐射问题上的失败，所以这也是整个经典物理学的“灾难”。

1900 年，普朗克利用内插法，把在短波和长波两种极限情况下成立的二式各自向中间部分延伸并联结起来，并加以数学变换，最后得到了著名的普朗克公式。

以后，有许多人做过实验，证明普朗克公式确实在可能测量到的所有波长和所有温度下都是成立的。得出了普朗克公式以后，普朗克就已经站到了量子论的大门口，下一步就是如何打开这扇大门了。普朗克经过紧张的八个星期的工作后，在德国物理学会的圣诞会（1900

年12月24日)上宣读了题为《关于正常光谱的能量分布定律》的论文。在这篇论文中他提出了具有重大意义的能量量子化的假设:

(1) 黑体的腔壁是由无数个带电的谐振子组成的,这些谐振子不断地吸收和辐射电磁波,与腔内的辐射场交换能量。

(2) 这些谐振子所具有的能量是分立的,它的能量与其振动频率 $\gamma$ 成正比,其关系可以写成 $\varepsilon=h\nu$,式中的 $h$ 即为普朗克常数。当谐振子与腔壁内的辐射场交换(吸收或辐射)能量时,也只改变 $\varepsilon$ 的整数倍,$\varepsilon$ 是谐振子能量的最小单位,即“能量子”。

普朗克在这篇论文中强调指出,“我们采取这种看法——并且这是整个计算中最主要的一点——认为 $E$(谐振子的能量)是由一些为数完全确定的、有限而又相等的能量子组成的。而对于这个有限而又相等的部分,我们应用了自然常数 $h=6.55\times10^{-27}$ 尔格·秒”。所以,人们通常把1900年12月24日作为量子论的诞生之日。

普朗克的能量量子化假设,大胆抛弃了经典物理学中物理量连续变化的旧观念,它不仅是对经典物理学的改造,而且是一次革命。它奠定了量子理论的基础,揭开了人类探索微观世界的序幕。普朗克能量量子化的概念,不仅与当时的物理学家们早已习惯了的方法相差甚远,而且与人们的普通常识也不相同,以至在量子假说提出后的最初五年时间里,并未引起物理学界的积极响应。德国在1908年出版的《自然科学与技术史手册》第二版,列举了1900年全世界120项发现与发明,就没有提到普朗克的发现。

### 5.9.3 经典力学的适用范围

首先,经典力学的应用受到物体运动速率的限制。当物体运动的速率接近于真空中的光速时,经典力学的许多观念将发生重大变化。如经典力学中,认为物体的质量不仅恒定不变,并且与物体的速度或能量无关。相对论的研究则表明,物体的质量将随着运动速率的增加而增大,物体的质量和能量之间存在着密切的联系。但当物体运动的速率远小于真空中的光速时,经典力学仍然适用。

其次,牛顿运动定律不适用于微观领域中物质结构和能量不连续的现象。19世纪和20世纪之交,物理学的三大发现,即 $X$ 射线的发现、电子的发现和放射性的发现,使物理学的研究由宏观领域进入微观领域,特别是20世纪初,量子力学的建立,出现了与经典观念不同的新观念。例如,量子力学的研究表明,微观粒子既表现为粒子性又表现为波动性,粒子的能量等物理量只能取分立的数值,粒子的速度和位置具有不确定性,粒子的状态只能用粒子在空间出现的概率来描述等。

但是量子力学的建立,并不是对经典力学的否定。对于宏观物体的运动,量子现象并不显著。在量子力学中可以证明,当粒子的能量比较大且作用粒子的力场变化比较缓慢时,量子力学的运动方程趋近于经典力学的规律。我们在量子力学和经典力学之间可以找到一个量,用它来标志在怎样的情况下可以运用经典力学和在怎样的情况下应该考虑用量子力学。这个量是普朗克常量 $h=6.626\times10^{-34}$[J·s],它具有[能量]×[时间]、[动量]×[长度]或[角动量]的量纲。如果表征粒子运动的量远大于普朗克常量,则量子现象可不考虑,即可应用经典力学;若表征粒子运动的量与普朗克常量可以比拟,则需要考虑用量子效应。

应当指出,现代物理学的建立和发展,并没有使经典力学失去存在的价值,经典力学仍将在它的适用范围内大放异彩。

# 学习资料

## 一、中国近现代著名科学家(物理学)简介

华罗庚(1910 年 11 月 12 日—1985 年 6 月 12 日),江苏金坛人,中国著名数学家,中国科学院院士,美国国家科学院外籍院士。他是中国解析数论、典型群、矩阵几何学、自守函数论与多元复变函数等很多方面研究的创始人与奠基者,也是中国在世界上最有影响的数学家之一,被列为芝加哥科学技术博物馆中当今世界 88 位数学伟人之一。



华罗庚出生于中国江苏金坛县,小学就读于金坛仁劬小学。1922 年,华罗庚进入金坛县立初级中学,1925 年夏毕业。由于家中贫穷,他只好到免学费的上海中华职业学校就读,但华罗庚还是因为家中无力提供杂费和住宿费而退学。1926 年,16 岁的华罗庚回到金坛后,帮助父亲料理杂货铺,同时开始自学数学。1929 年 12 月,华罗庚在《科学》杂志第 14 卷第 14 期上发表《Sturm 氏定理之研究》。1930 年 12 月,华罗庚在《科学》第 15 卷第 2 期上发表《苏家驹之代数的五次方程式解法不能成立之理由》,被清华大学数学系主任熊庆来教授发现,邀请他来清华大学。从 1931 年起,华罗庚在清华大学边学习边工作,用一年半时间学完了数学系的全部课程,并同时自学了英、法、德语,在国际学术杂志上发表了三篇论文,被熊庆来破格任用为助教。1936 年赴英国剑桥大学访问、学习。1938 年回国后任西南联合大学教授。1946 年赴美国,任普林斯顿数学研究所研究员、普林斯顿大学教授,1950 年回国。1955 年被选聘为中国科学院学部委员(院士)。1958 年主持创立中国科技大学数学系,并任中国科技大学副校长兼数学系主任。1979 年加入中国共产党。1985 年 6 月 12 日,华罗庚应邀到日本东京作学术报告,报告结束后突发心脏病倒在讲台上,送院后证实不治,终年 75 岁。

数学是自然界的普适语言;某种意义上,数学是物理学的表征,物理学是数学的实证模型——只要数学方程是自洽的,则必然有物理的自然客观模型与之对应——尚未发现而已!

物理学是当代技术之源——需要数学进行标准化。

当代最伟大的数学家之一陈省身教授,极力赞成“物理”与“数学”是一家。他曾在一首诗中写道:

物理几何是一家,共同携手到天涯。<br>
黑洞单极穷奥秘,纤维联络织锦霞。<br>
进化方程孤立异,对偶曲率瞬息空。<br>
筹算竟有天人用,拈花一笑不言中。

## 二、诺贝尔物理学奖简介

**1931 年未颁奖**。

1932 年诺贝尔物理学奖。因创立量子力学和应用该理论发现氢的同素异形体。获奖者:海森伯(德国)。评价:作为量子力学重要的奠基人之一,海森伯因矩阵力学和他的不确定性原理而闻名于世。虽然不确定原理的证明并不复杂,但他却在挑战人的逻辑的最深

层次。不确定原理告诉我们,不可能同时准确测出一个粒子的动量和位置,它们的误差乘积为一个常数。那么一个粒子是否存在一个准确的动量和位置?海森伯说,“如果人们要弄明白‘一个物体的位置’,例如一个电子的位置这个说法是什么意思,就必须指定一个用以测量‘电子位置’的实验,否则这个说法就没有任何意义。”

**1933年诺贝尔物理学奖**。发现原子理论新的有效形式。获奖者:E. 薛定谔(奥地利)、P. A. M. 狄拉克(英国)。评价:这是另两位量子力学的宗师。薛定谔提出的波动力学和海森伯的矩阵力学是量子力学殊途同归的两个形式(数学上等价)。至今薛定谔方程仍是量子力学的基本方程(海森伯的方程难以理解目前大学教授量子力学基本用薛定谔的方程)。而狄拉克导出了薛定谔方程的相对论形式,即狄拉克方程,并预言存在反电子(电子带正电)。此外狄拉克创造出了狄拉克符号,差不多是在0的位置定义一个1除以0(虽然用的是极限的方法回避了矛盾,但我觉得实际上就是1除以0),对很多学科(自控、电子)都有影响。

**1934年未颁奖**。

**1935年诺贝尔物理学奖**。发现中子。获奖者:查德威克(英国)。评价:自从卢瑟福发现质子(卢瑟福是非常著名的物理学家发现了质子,可惜拿的是诺贝尔化学奖)以来,人们发现原子核的质量和质子质量不相等,这就表示原子核里还有其他的结构。他预言了中子的存在。后来很多物理学家的研究都到了发现中子的边缘,包括居里夫人的女儿艾伦娜·居里,但都放弃了。

查德威克1891年出生在英国柴郡,曼彻斯特维多利亚大学毕业。中学时代并未显现出过人天赋。他沉默寡言,成绩平平,但坚持自己的信条:会做则必须做对,一丝不苟;不会做又没弄懂,绝不下笔。因此他有时不能按期完成物理作业。而正是他这种不慕虚荣、实事求是、“驽马十驾,功在不舍”的精神,使他在科学研究事业中受益一生,在前人的路上发现了中子。中子的发现对认识原子核内部结构是一个转折点,具有重大的理论意义。在中子未被发现以前,人们对于原子核的内部结构不完全清楚,发现中子之后,人们才知道原子核是由中子与质子组成的。而且它对原子核的反应有重要的作用。因此,一般认为,人类进入原子能时代的大门是被中子敲开的。

**1936年诺贝尔物理学奖**。发现宇宙射线;发现正电子。获奖者:V. F 赫斯(奥地利)、C. D. 安德孙(美国)。评价:宇宙射线的发现是非常偶然的。赫斯发现某种射线的强度在白天和夜间测量结果相同,因此赫斯断定这种射线不是来源于太阳的照射,而是宇宙空间。最初这种辐射被称为“赫斯辐射”,后来被正式命名为“宇宙射线”。现在知道,宇宙射线就是能量很高的各种粒子。而正电子的存在是由于狄拉克方程计算的结果中存在一组负能量解(即存在负能级),所以狄拉克预言存在正电子充满负能级,使电子(负)无法跃迁到负能级里去。后来安德孙用威尔孙云室来发现了正电子,证明了狄拉克的预言。

**1937年诺贝尔物理学奖**。通过实验发现受电子照射的晶体中的干涉现象。获奖者:C. J. 戴维孙(美国)、汤姆孙(英国)。评价:自德布罗意提出物质波以来,关于物质波到底是什么形式众说纷纭。人们迫切想通过实验来观察到物质波的存在,而当时已知的动量最小的物质莫过电子(动量小那么波长就长一些波动性就容易表现,比如人的波长实在太短所以无法表现),这两人通过晶体间距作缝隙(这就多亏前面布喇格的贡献了)观察到了晶体的衍射现象从而无可辩驳地证明了物质波的存在。

**1938年诺贝尔物理学奖**。用中子产生新的放射性元素和开展慢中子核反应的研究工作。获奖者:费米(意大利,后加入美籍)。评价:古代人梦想的炼金术首先被卢瑟福实现,

他用高速 α 粒子(即氦核)轰击氮核,得到了氧核。这样人们实现了不同元素的转化。理论上有了点石成金的可能。但由于 α 粒子带正电,所以他不能作用于原子序数高的元素(即带的正电多,同性相斥)。而费米通过用慢中子反应实现了铀的裂变,为制取原子弹打下了基础。他也是人类第一座核反应堆的建立者,被称为中子物理学之父,不过诺贝尔奖这次摆了一个乌龙,当费米做完铀核反应时以为自己得到的是一种新的 93 号元素。不过哈恩(1944 年诺贝尔化学奖得主)却证明铀在反应时分成了两种已知元素,也就是诺贝尔奖搞错了。不过费米不愧有大科学家风范,此时此刻,费米考虑的不是个人的名誉得失,他在别人成就的基础上继续向前迈进。在裂变理论的基础上,费米很快提出一种链式反应假说,为打开原子弹这个魔盒迈出了关键一步。

**1939 年诺贝尔物理学奖**。发明和制造回旋加速器并用其产生人工放射性元素。获奖者:劳伦斯(美国)。评价:加速器对于研究物理有非常重要的意义,比如制造中子和同位素。很多物理现象就是通过加速器加到一个高速度才可能实现。现在如火如荼的欧洲强子对撞机很大一部分就是加速器。而回旋加速器相对于直线加速器占地面积成本要小很多,所以在当时甚至现在起到了非常重要的作用。不过回旋加速器的最大可加的速度有其极限,所以现在要想取得高速还得靠直线加速器。

**1940 年未颁奖**。

## 思考与练习

1. 如何看待牛顿力学的局限性。
2. 惯性力是真实的力吗?
3. 参考系的本质特征是什么?
4. 惯性定律就是牛顿第一定律吗? 为什么?
5. 牛顿第一定律是牛顿第二定律的特例吗?
6. 为什么说牛顿第三定律不是普遍成立的?
7. 简谐振动中位移共振与速度共振有什么区别与联系?
8. 为什么弹性介质波的动能与势能是同步的?

# 第6章

# 热力学系统与熵

我们知道，热力学是热学的宏观理论（热学的微观理论的普通物理学的层次是分子动理论，理论物理层次是统计物理学），是根据热运动的实验结果，进行归纳、综合从而形成的系统理论。它不涉及物质的微观结构和微观粒子的相互作用，也不涉及特殊物质的具体性质，是一种唯象的宏观理论。热力学的任务是：研究热运动的规律，研究与热运动有关的物性及宏观物质系统的演化。通过对热现象的观测、实验和分析，人们总结出热现象的基本规律：热力学第零定律、热力学第一定律、第二定律和第三定律。这些基本规律是对无数经验的总结，适用于一切宏观物质系统，以这些基本规律为基础，应用数学方法，通过逻辑演绎可以得出物质各种宏观性质之间的关系、宏观过程进行的方向和限度等结论，只要其中不加上其他假设，这些结论就具有同样的可靠性和普遍性。

热力学定律以及三个基本状态函数温度、内能和熵构成了完整的热力学理论系统。为了在各种不同条件下讨论系统状态的热力学特性，还引入了一些辅助的态函数，如焓、亥姆霍兹函数（自由能）、吉布斯函数等。从热力学的基本定律出发，应用这些态函数，经过数学推演得到系统平衡态的各种特性的相互联系，这就是热力学的方法，也是热力学的基本内容。热力学理论是普遍性的理论，对一切物质都适用。在涉及某种特殊物质的具体性质时，需要把热力学的一般关系与相应的特殊规律结合起来。例如讨论理想气体时，需要利用理想气体的状态方程等。平衡态的热力学理论已经相当完善，并且得到了广泛的应用。然而在自然界中，处于非平衡态的热力学系统（物理的，化学的，生物的）和不可逆的热力学过程是大量存在的，并且和许多重要现象有关。因此，非平衡态热力学和不可逆过程热力学是正在发展的一个重要领域。此外，热力学理论不考虑物质的微观结构，把物质看作连续体，用连续函数描写物质的性质，因此不能解释涨落现象，这是热力学的局限性。

本章从普通物理学的层次——热学和分子动理论两个方面讨论了热学系统的基本概念和基本原理，建立和发展的过程，详细讨论了热力学定律的物理含义、相互关系及其历史地位。这些讨论对于深入理解热力学的概念、原理具有现实的意义。

## 6.1　热力学系统的基本概念

热力学研究的对象是由大量微观粒子（分子或其他粒子）组成的有限的宏观物质系统，热力学是热运动（热运动是指大量微观粒子的随机运动）的宏观理论，从这样的前提出发研究物质的性质和运动规律。热力学的完整理论系统是由三个基本定律以及相应的三个基本状态（函数温度、内能和熵）以及一些辅助的态函数构成。首先，我们将在这一节中简单地介绍热力学的一些基本概念。

### 6.1.1 热学发展史简介

人类很早就对热有所认识，并加以应用。但是将热力学当成一门科学且有定量的研究，则是由 17 世纪末开始的，也就是在温度计制造的技术成熟以后，才真正开始了对热力学的研究。热机的发明和应用，对热学理论的发展起到了巨大推动和促进作用。

热力学发展史，基本上就是热力学与统计力学的发展史，大致可分成四个阶段。

第一个阶段：17 世纪末到 19 世纪中叶。此时期累积了大量的实验与观察的结果，并制造出蒸汽机，对于“热”的本质展开研究与争论，为热力学的理论建立作好了预热。在 19 世纪前半叶，首先出现了卡诺理论，热机理论(第二定律的前身)和功热互换的原理(第一定律的基础)。这一阶段的热力学还留在描述热力学的现象上，并未引进任何的数学算式。

第二个阶段：19 世纪中叶到 19 世纪 70 年代末。此阶段热力学的第一定律和第二定律已完全理论化。基于功热互换原理建立了热力学第一定律，由第一定律和卡诺理论的结合，促使热力学第二定律的成熟。另一方面，以牛顿力学为基础的气体动力论也开始发展，但这时期人们并不了解热力学与气体动力论之间的关联。

第三个阶段：19 世纪 70 年末到 20 世纪初。这个时间内，首先由玻耳兹曼将热力学与分子动力学的理论结合，而导致统计热力学的诞生，同时他也提出非平衡态的理论基础，至 20 世纪初吉布斯提出系综理论建立统计力学的基础。

第四个阶段：20 世纪 30 年代到今。主要是量子力学的引进而建立了量力统计力学，同时非平衡态理论更进一步的发展，形成了近代理论与实验物理学中最重要的一环。

### 6.1.2 热力学与统计物理学

热力学是热学的宏观理论，是根据实验结果综合整理而成的系统理论，它不涉及物质的微观结构和微观粒子的相互作用，也不涉及特殊物质的具体性质，是一种唯象的宏观理论，具有高度的可靠性和普遍性。

统计物理学是热运动的微观理论。物体是由大量的原子或分子组成的。这些粒子的任何一种排布都是该物体的一种状态。一个孤立的平衡态系统，出现任何一种微观状态的概率相等，称为等概率原理。统计物理学就是从这一基本原理出发，对所研究的系统给出简化模型，通过统计学方法，进行逻辑演绎和理论计算，导出大量微观粒子组成的物体的热运动规律。如果粒子遵从经典力学的运动规律，对粒子运动状态的描述称为经典描述；如果粒子遵从量子力学的运动规律，对粒子运动状态的描述称为量子描述。

统计物理不追求个别粒子的运动细节，而是研究集体行为表现的规律——统计规律性。量子统计和经典统计的主要区别在于对微观运动状态的描述上，而在统计原理上是相同的。经典统计认为微观粒子是可区分的，而量子统计认为微观粒子是全同的。

热力学理论总结出来的经验定律不考虑物质的具体微观结构，所以有高度的普遍性，适用于一切物质；所研究的系统是一个连续体，用连续函数来描写，讨论宏观量之间的关系，不深入讨论现象的本质；主要研究平衡态和可逆过程；对于非平衡态和不可逆过程只给出定性的和方向性的估计。热力学不能解决涨落问题。

统计物理学是从物质的微观结构出发，对微观量求统计平均值。这个统计平均值就是热力学的宏观物理量。所以，统计物理方法建立了物体的微观状态和宏观状态之间的关系，讨论了宏观量的本质，解释了涨落现象和从非平衡态到平衡态的不可逆过程。当粒子数很大时，涨落很小，热力学和统计物理学的结果是一致的。但当系统的尺寸很小，粒子数较少

时,涨落现象显著,或在相变临界点附近,热力学理论将不适用。所以,统计物理学是理论更严格、更抽象、适用范围更广泛的学科。既不要怀疑热力学的可靠性,它已经历了200多年的科学检验,很多热力学参量是可以直接测量的;也不要迷信统计物理学的万能性,因为从微观统计计算宏观特性常常遇到很大的困难。

### 6.1.3 热力学物质形态的基本概念

**1. 热力学系统**

热力学系统,是指大量微观粒子(分子或其他粒子)组成的有限宏观物质系统。

外界,与系统发生相互作用的其他物体,即与热力学系统相互作用的环境。

间壁,包围热力学系统的一层外界物质。

透热壁,能使热力学系统与外界进行热传递作用的间壁。

绝热壁,不能使热力学系统与外界进行热传递作用的间壁,是理想模型。

孤立系统,忽略与外界任何相互作用的系统,系统与环境之间既无物质的交换,也无能量的交换,是理想模型。

封闭系统,与外界只有能量交换,忽略物质交换的系统。系统与环境之间没有物质的交换,只有能量的交换,是理想模型。

开放系统,与外界有能量交换和物质交换的系统,系统与环境之间既有物质的交换,也有能量的交换。

**2. 热力学物态**

(1) 理想气体:在宏观上,理想气体是一种在任何情况下都遵守玻意耳定律、盖-吕萨克定律和查理定律的气体。

理想气体的微观模型具有三个基本特征:其一,分子与容器壁和分子与分子之间只有在碰撞的瞬间才有相互作用,其他时候的相互作用可以忽略不计;其二,分子本身的体积在气体中可以忽略不计,即对分子可采用质点模型;其三,分子与容器壁以及分子与分子之间的碰撞属于牛顿力学中的完全弹性碰撞,即忽略分子大小和形状,将分子看作质点;除碰撞外,忽略分子间作用的气体,被认为是理想气体。

实验证明,实际气体中分子本身的体积约占气体体积的千分之一,在气体中分子之间的平均距离远大于分子的几何尺寸,所以将分子看成质点是完全合理的。从另一个方面看,对已达到平衡态的气体如果没有外界影响,其温度、压强等态参量都不会因分子与容器壁以及分子与分子之间的碰撞而发生改变,气体分子的速度分布也保持不变,因而分子与容器壁以及分子与分子之间的碰撞是完全弹性碰撞也是理所当然的。

所以,经过抽象与简化,理想气体可以看成是一群彼此间无相互作用的无规则运动的弹性质点的集合,这就是理想气体的微观模型。

(2) 范德瓦尔斯气体:考虑到理想气体把分子看作没有大小的质点所带来的误差而引入了修正系数 $b$、考虑到理想气体忽略了分子之间的相互作用力而引入了修正数 $a$,修正系数 $b$ 和 $a$ 来修正理想气体的状态方程 $pV=RT$,得到了关于实际气体的范德瓦尔斯方程

$$\left(p+\frac{a}{V^2}\right)(V-b)=RT \tag{6-1-1}$$

但是,由于这种建立在气体微观结构基础上的修正也是对气体微观结构的一种近似,它与理想气体模型的区别仅在于近似程度不同。正因为这一点,人们将遵守范德瓦尔斯方程的气体称为"范德瓦尔斯气体"。

(3) 实际气体：实际存在的气体，又称真实气体。不能忽略分子本身占有的体积和分子间的相互作用力等因素的气体。严格地说，热机中遇到的各种气态工质都是实际气体，不能用理想气体状态方程来描述。只是在一定条件下，有些气体（如大气中的空气、锅炉里的蒸汽、内燃机和燃气轮机中的燃气等）可按理想气体作近似处理；而另外一些气体（如压力比较高而温度接近其液化温度的水蒸气）的性质偏离理想气体较远，必须按实际气体处理。

(4) 晶体：晶体又可以分为单晶体和多晶体。单晶体是指有规则的几何形状、各向异性和有固定熔点的晶体，例如冰盐、石英和云母等。多晶体是指金属晶体，其特性有各向同性和固定熔点等。

(5) 非晶体：没有固定熔点，连续相变，各向同性。例如，玻璃，沥青，塑料等。

(6) 液晶体，简称液晶，具有确定的体积、形状，没有确定的熔点、凝固点等。

(7) 液体：物质存在的一种状态，有比较固定的体积而没有固定的形状，可流动。增温或减压一般能使液体汽化，成为气体。加压或降温一般能使液体固化，成为固体。

**3. 热运动的定义**

第一种定义：热运动是微观粒子永不停息的无规则运动——是微观粒子的一种运动形式。

第二种定义：物质内部大量微观粒子无规则运动的整体表现为宏观物质的热运动——是宏观物质的一种运动形式。

第二种定义的优点：

(1) 热学的研究对象——大量微观粒子构成的宏观物质；

(2) 适合学生的认知特点——首先认识宏观物质——然后再认识微观粒子；

(3) 微观粒子的运动不仅仅只有无规则运动，还有更复杂更高级的运动形式。

**4. 热力学变量**

热力学变量是描述热力学系统性质的，用于描述系统的热力学状态及其变化。比如，体积 $V$、压强 $p$、温度 $T$、输运系数 $\eta$ 等。与力学量的分类相似，热力学变量可以分为：

(1) 广度性质(extensive properties)：其数值大小与系统所含物质的量成正比，如质量 $m$、体积 $V$、内能 $U$ 等。

(2) 强度性质(intensive properties)：其数值大小与系统所含物质的量无关，如温度 $T$、压强 $p$、输运系数 $\eta$ 等。

**5. 热力学平衡态**

热力学平衡态的类型有：

热平衡(thermal equilibrium)：系统中温度处处相等。

力平衡(force equilibrium)：系统各部分之间，没有不平衡的力存在。

相平衡(phase equilibrium)：系统中各相的组成和数量不随时间而变。

化学平衡(chemical equilibrium)：系统中发生的化学反应达到平衡，系统中各系统的组成不随时间而变。

**6. 状态函数与状态方程**

状态函数是描述系统状态性质的函数。系统的特征变化，系统的状态和性质也变化，是单值函数。其改变值只取决于始、终态，与过程历史无关，其微小变化在数学上是全微分。如温度、内能和熵焓、亥姆霍兹函数(自由能)、吉布斯函数等都是态函数。

状态方程：系统状态函数之间的定量关系式，如 $pV=nRT$。

**7. 过程和途径**

过程：系统所发生的一切变化。热力学系统的过程可以分为：

(1) 简单状态变化过程(无相变、化学反应发生)：等温过程(isothermal process)：$T_{始}=T_{终}=T_{环}$；等压过程(isobaric process)：$p_{始}=p_{终}=p_{环}$；等容过程(isochoric process)：$V_{始}=V_{终}$；绝热过程(adiabatic process)：$Q=0$。

(2) 相变过程：在均匀单相内，出现了不同成分或不同结构(包括原子、离子或电子位置及位向的改变)、不同组织形态或不同性质的相的过程。

(3) 化学变化过程：生成了新物质的变化过程，还发生能量的变化，这种变化常以放热、发光等的形式表现出来。

途径：系统同一始态到达同一终态的不同方式。

**8. 热和功**

系统与环境之间进行能量交换的两种方式：热传递和做功。系统与环境之间发生热传递时规定：系统吸热 $Q(+)$；放热 $Q(-)$；做功：除了热以外的各种能量交换的形式。

(1) 膨胀功，又称体积功：

$$\delta W_{体}=p\,dV \tag{6-1-2}$$

(2) 非体积功：又称有用功，其他功。$W'$，包括电功、机械功、表面功，黏附功等。一般规定系统对环境做功 $W$ 为正，环境对系统做功 $W$ 为负。

$$\delta W=\delta W_{体}+\delta W'=p\,dV+\delta W' \tag{6-1-3}$$

热和功都是能量传递的形式，不是系统固有的性质，不是状态函数，它们的变化与具体过程有关。

### 6.1.4 热力学的基本原理及概念

热力学是热运动的宏观理论。其研究方法的特点是，不考虑物质的微观结构，从大量的实践和实验总结出来的基本规律出发，演绎推导出宏观物理量之间的关系及其变化规律。热力学将所研究的对象用一组热力学变量来描述其状态。

**1. 热力学第零定律**

(1) 定律表述：两个物体同时与第三个物体达到热平衡，这两个物体也处于热平衡。

(2) 定律的物理意义：热力学第零定律是通过实验推论对经典概念“温度”下定义的：无论多少个物体互相接触都能达到热平衡，并且如果 A 物体同时与 B、C 两物体处于平衡态，则 B、C 两物体接触时也一定处于平衡态而不发生新的变化，这一热平衡规律称为热力学第零定律。

(3) 概念：由此可以引入一个状态函数——温度 $T=T(p,V,\sigma,\cdots)$。温度是判定一系统是否与其他系统互为热平衡的标志。关于温度的进一步讨论，读者可以参阅本 6.2 节和 3.6 节。

**2. 热力学第一定律**

(1) 定律表述

文字表述：能量守恒和转化定律。

定量表述Ⅰ：

$$dQ=dW+dU \tag{6-1-4}$$

式中 $dQ$，$dW$ 和 $dU$ 分别是系统与外界交换能量(热量)，做功以及系统的内能改变。

定量表述Ⅱ：

$$TdS = dU + pdV \tag{6-1-5}$$

式中 $S$ 是系统的熵，$U$ 是系统的内能，$p$ 是系统的压强，$V$ 是系统的体积。

(2) 定律的物理意义：热力学第一定律给出了各种形式的能量在相互转化过程中必须满足的能量守恒，对过程进行的方向并没有给出任何限制。然而，在自然界中，任何宏观自发过程都是具有方向性的。自发过程是指在不受外来干预的条件下所进行的过程。在热力学中，所谓过程的方向，总是指自发过程的方向。因为一旦加上了外来的干预，过程的方向问题就变得毫无意义了。

(3) 概念：热力学第一定律就是能量守恒定律，是后者在一切涉及热现象的宏观过程中的具体表现。描述系统热运动能量的状态函数是内能。通过做功、传热，系统与外界交换能量，从而改变系统的内能。

**3. 热力学第二定律**

(1) 定律表述

克劳修斯(Rudolf Julius Enmanvel Clausius，1822—1888)表述：不可能把热量从低温物体传到高温物体而不引起其他影响。

开尔文(Lord Kelvin，1824—1907)表述：不可能制造一种机器，它能从一个低温度热源吸取热量使之完全转变为功而不产生其他影响。

定量表述：孤立系统的熵不可能减少。熵是个态函数，其定义式为

$$S = \int \frac{dQ}{T} \tag{6-1-6}$$

式中 $dQ$，$T$ 分别是热量的微分和绝对温度；绝对温度和摄氏温度的关系为：$T=t+273.15\text{K}$，这里 273.15K 为水的冰点的绝对温度。

(2) 定律的宏观物理意义

热力学第二定律的实质在于它指出了，自然界中一切与热现象有关的实际宏观过程都是不可逆的；或者说，一切与热现象有关的自发实际宏观过程都是有方向的；还可以说，热力学第二定律指出一切涉及热现象的宏观过程中能量转换或传递的方向、条件和限度。热力学第二定律所揭示的这一客观规律，向人们指出了实际宏观过程进行的条件和方向。在孤立系统内，热力学第二定律具有深刻的物理内涵，该定律揭示了时间的单向性，说明了物理因果关系的绝对性等。

(3) 热力学第二定律的统计意义——定律的微观物理意义

① 熵的定义——玻耳兹曼公式

单原子气体分子在空间的各种布居方式有 $W_{sp}$ 种，又独立于空间布居而在各速度范围内的不同配置方式有 $W_v$ 种，于是该气体系统的热力学概率为：$W=W_{sp}W_v$。若气体是多原子气体，还得考虑分子的转动和振动，情况要复杂些。但是，这个热力学概率不具有可加性(一般来说，处理具有可加性的量总是比较简便的)，于是取对数，得到：$\ln W=\ln W_{sp}+\ln W_v$。

对于热力学概率为 $W$ 的宏观态，定义一个物理量

$$S = k\ln W \tag{6-1-7}$$

称为熵，它指明了热力学过程进行的方向，是相应的态函数，熵的大小反映了系统所处状态的稳定性。熵具有可加性的广延量。

② 热力学第二定律的统计意义

从微观角度看，不可逆过程是从热力学概率小的宏观态向热力学概率大的宏观态的过

渡,概率的法则决定了这样的不可逆过程是可以自发地进行的。当系统达到平衡态时其热力学概率达到最大值,其宏观分布为最概然分布。由于平衡态热力学概率比与其偏离的非平衡态的热力学概率大得非常多,所以概率法则决定了不可逆过程的逆过程发生的概率非常小,实质上不可能发生。这就是热力学第二定律的统计意义。下面再分析"功变热"及"热传导"的不可逆性。

功变热,实质上是机械能变为内能的过程。机械能,是所有的微观粒子都作同样的定向运动所对应的一种有序能量;而内能,则是与微观粒子无规则热运动联系着的无序能量。与有序运动对应的机械能,极易通过耗散的因素变为与无序运动对应的内能,并被均分到各个运动方向及其自由度上去,但相反的过程却只有非常小的概率发生,实际上根本不会实现。

在热传导的初始状态,总能区分出物体系统温度不同的各部分。虽然整个系统的微观粒子都处于无规则的热运动之中,但毕竟可以根据不同部分温度的高低,指出哪里的微观粒子热运动平均能量最大、运动最无序,哪里次之。这比起热传导的终了状态,不再能区分各处微观粒子的热运动激烈程度,显然初态较为有序。因此,热传导过程的方向,就是大量分子热运动更加无序化的发展方向。

**4. 热力学第三定律**

1906年,美国科学家能斯特(1864—1941)在研究低温下各种化学反应的性质时,通过总结大量实验资料得出了一个普遍规则,即系统的熵在等温过程中的改变,随着绝对温度的趋近于零而趋于零。

(1) 定律表述:

一般表述:绝对零度不可能达到。

能斯特表述:当温度为绝对零度时,任何物质的熵都相等。

$$\lim_{T\to 0\mathrm{K}}(\Delta S)_T=0 \tag{6-1-8}$$

此式称为能斯特定理,一般情况下可作为热力学第三定律的一种表述。

(2) 定律的物理意义:普遍而言,热力学第三定律可以用绝对零度不能达到原理来表述,即不可能施行有限的过程把一个物体冷却到绝对零度。

整个热力学就是从这些基本定律出发,应用数学方法、逻辑演绎导出描写宏观物体性质的参量之间的关系、运动过程的方向和限度等的结论。

## 6.2 热力学温度及其物理意义

温度是热学中最重要的基本概念之一,也是国际单位制七个基本量中在热学领域内特有的物理量。温度概念的建立和测量是建立在热平衡基础之上的。从定性的或宏观的意义上说,温度是表示物体冷热程度的物理量;而从微观的意义上说,温度是物体内部分子热运动平均动能的标志,温度越高,物体内部分子的热运动越剧烈。在自然界中任何物体的物理和化学性质都与温度有密切的关系,例如春、夏、秋、冬四季,物质的汽态、液态、固态都和温度的高低有直接关系。人类以及动物的生活规律和生活习性也是与温度息息相关的。温度这个量也是工业生产中的一个重要参量,温度的测量与人们的学习和生活有着密切的联系,许多工业部门都要精确地测量温度,准确地控制温度。如果温度测量不准或控温系统失灵,就会影响生产,甚至造成事故。所以与温度有关的很多问题是现代物理学研究的热点,对温

度这个量的科学认识和掌握，对工农业生产、科学实验和物理学习，具有现实意义。

本节我们从温度的宏观和微观定义入手，探讨在热力学中温度的物理意义，然后，重点讨论负绝对温度的定义、实验基础及其物理意义。

### 6.2.1　温度的宏观定义

温度的概念来源于人们对物体冷热程度的感觉，在日常生活中常见的一些温度现象是和热的传导联系在一起的。但对温度的特性和本质的认识，就不能仅靠感觉来定义，更不能靠它来作比较温度的标准。因此我们需要科学的温度的定义。

在生活、生产和技术中，更需要对物体冷热程度的精确描述——定量化描述，需要引入新的物理量——温度。在初中物理教材中，温度的定义为表示物体的冷热程度的物理量，而高中物理教材定义温度为表示物体热运动程度的物理量。在大学物理学中，温度概念的建立和定量测量都是以热平衡现象为基础的。所以，在宏观热力学中，温度是建立在热平衡定律(即热力学第零定律)基础上的一个概念。能传热的两个物体相互接触称为热接触。不受外界影响，两热接触的物体达到冷热程度相同时称两物体互为热平衡。经验告诉我们，两个不相接触的物体 A 和 B，如果分别和第三个物体 C 互成热平衡，则物体 A 和 B 也处于热平衡。这个事实作为建立“温度”的实验基础。热力学称“热平衡定律”。为了热力学体系的完整性，物理学家称之为热力学第零定律。

由热力学第零定律可知，一切互为热平衡的物体都具有某一相同的宏观物理性质，表征这一宏观物理性质的量就称为温度。这就是温度这个态函数的实验基础和理论基础。

### 6.2.2　温度的微观定义

在微观分子物理学中，通过研究，发现物体的冷热程度完全由大量分子无规则运动的强度所决定，分子的平均平动能是表征分子无规则运动强度最适当的物理量。平均平动能愈高，分子无规则运动就愈剧烈，物体也就愈热。分子无规则运动的平均平动能所具有的特性正和我们在宏观理论中所引进的温度性质相当。因此，温度的概念在分子运动理论中获得了新的意义：温度是大量分子作无规则运动的平均动能的标志。温度与分子无规则运动平均平动动能的具体关系是：$E_k=3kT/2$，式中 $E_k$ 表示物体内大量分子的平均平动动能，$k=1.38\times10^{-23}$ J/K，称为玻耳兹曼常数，$T$ 是热力学温度。

应当注意的是：因为平均平动动能是一个对大量分子而言的统计平均值，所以温度只有统计意义，对于个别分子，说它有多少温度是没有意义的。另外，只有当物体处于热平衡时，才能定义它的温度。

从微观来看，物体的温度与分子热运动密切相关。在分子运动论中，从气体的压强分子公式中可以看到，温度是分子热运动平均动能的标志。微观物质的相互作用，在固体和固体之间或固体与气体之间，物质分子间的相互作用表现为分子热运动过程中的完全弹性碰撞。在一维质心弹性碰撞中，动量守恒。两弹性物体的动量数值相等或质量相等时，动能守恒。

如果采用气体分子的平均动能作为能量定义的标志，则在固体分子和气体分子的弹性碰撞中须满足对物质温度的描述。在温度为一定值、固体分子和气体分子的质量相差很大时，发生在固体分子和气体分子之间的碰撞在宏观统计下，不能满足于固体分子，气体分子原平均速度在碰撞后不变的描述。因此，采用平均动能对温度定义不能满足于对物质温度的描述。

如果采用物质分子的平均动量来对温度定义，则只要定义物质的温度同物质分子的平

均动量成正比,在分子间的完全弹性碰撞中,就能满足对平衡状态的物质温度进行描述。

根据以上关于温度定义的讨论,我们可以有下面的一些思考:

(1) 采用物质分子的平均动量来对温度定义,物体在气体中运动时和气体有一相对的运动速度,气体分子和物体分子相互碰撞间必然有一相对动量,其值大于气体分子和物体分子在其温度下的动量。因此在它们之间的碰撞也必然有一相对的温度。

(2) 由于气体分子的运动速度很高。而日常生活中物体的运动速度很低,物体相对于气体分子的相对运动忽略不计,当物体的运动速度很高时,如流星,我们就不能忽略不计了,物体分子与气体分子间的相对温度可以发出能量较高的可见光。

(3) 在地球表面的空间中充满了空气,地球的引力对气体分子在其自由程内沿竖直的方向有一加速的过程。这样,在地球大气的不同层面上,气体分子存在速度差异,也就存在着一定温度差异。这是由地球引力引起的。地球引力引起的这种温差现象,随距地表的高度增加,温度变化会逐渐明显。如果大气的厚度存在变化,那么地表的温度也会存在变化。

(4) 如果大气的平均分子量增加时,地表的温度也会增加。由于温度与动量成正比,气体分子自由程内,分子质量越大,由分子获得的动量越大。

温度像长度、质量和时间一样,在国际单位制(SI)中被列为基本物理量。关于温度的测量方法以及温标的定义请读者参阅本书3.6节。

### 6.2.3 负绝对温度的概念

如果我们能把全部热量从一个物体中抽出来,那么就会达到绝对零度(0K),由于温度不会再“下降”,物体不会变得更冷了。反之,如果把热量一直不断地加给这个物体,那么它的温度能够无限地增加。这样看来,绝对温度似乎界于零度和正无穷大之间的范围内。这是我们通常用普通温度计测量出来的那种温度。按照经典物理学的推论,似乎任何系统不可能达到绝对零度以下的负的绝对温度,但是有些物理系统却可以具有负的绝对温度。

负绝对温度(negative absolute temperature)现象是美国科学家潘德(Pound)和珀塞尔(Purcell)等人于1951年在氟化锂(LiF)单晶的核自旋系统的研究中最早发现的。他们用很强的磁场使氟化锂晶体中的锂核磁矩的方向绝大部分沿磁场方向排列,因而处于低能的正温度状态。然后突然使磁场反向,但锂核磁矩却不能随着磁场同步反向,造成与磁场方向相反,使系统绝大多数锂核处于高能的负温状态。自此,负温问题即成为统计物理中的一个重要课题。1956年,拉姆赛(Ramsay)进行了系统的研究和整理,形成了完整的理论。

下面从经典热力学绝对温度的统计定义出发,讨论负绝对温度的含义及其物理本质。

**1. 绝对温度的经典统计定义**

由分子动理论可以得到绝对温度的经典统计定义,即微观定义。

从微观来看,物体的温度与分子热运动密切相关。温度与分子无规则运动平均平动动能的具体关系是:$\bar{\omega}=\frac{1}{2}m\overline{v^2}=\frac{3}{2}kT$。由此可得绝对温度的一个定义式

$$T=\frac{2}{3}\times\frac{\bar{\omega}}{k} \tag{6-2-1}$$

式中 $\bar{\omega}$ 表示物体内大量分子的平均平动动能,$k=1.38\times10^{-23}$K/J,称为玻耳兹曼常数,$T$ 是热力学温度,单位是开尔文。在这里,温度反映了系统内部分子无规则运动的剧烈程度,即温度是分子平均平动动能的标志。因为平均平动动能是对大量分子而言的统计平均值,所以,经典物理学的温度只有统计意义,对于个别分子,说它有多少温度是没有意义的。同

时，只有当物体处于热平衡时，才可以定义它的温度。

显然，这个定义的缺点在于假设粒子(分子)的能量是连续分布的，由它得出的结论与新的实验事实不符。例如，由式(6-2-1)，当 $T\to0$ 时，则必 $\bar{\omega}\to0$。而量子力学指出，粒子能量(或其运动状态)是量子化的。在 $T\to0$ 时，粒子只是处于最低能级，仍有所谓“零点能”存在，这是实物粒子具有波粒二象性的必然结论，它已经被实验证实。即经典统计关于温度的微观定义与已经为实验证实的“零点能”存在不符。

与此同时，因为平均平动动能 $\bar{\omega}$ 恒为正值，所以由式(6-2-1)得到的只能是正绝对温度 $T$，而负绝对温度是不存在的。

**2. 绝对温度的量子统计定义与负绝对温度的引入**

由玻耳兹曼能量分布律可以得到绝对温度的另一种微观定义。

量子力学指出，粒子只能处于能量不连续的不同能级，即能级是量子化的。系统的温度越高，分布在高能级上的粒子数越多。因此，温度的微观定义可以较明确地表达为：温度是表征粒子数按能级分布情况的物理量。玻耳兹曼能量分布为 $n=n_0\mathrm{e}^{-\frac{\varepsilon}{kT}}$，式中 $n_0$ 是具有初始能量为 $E_0$ 的粒子数，$n$ 具有能量为 $E$ 的粒子数。由此式可得绝对温度的又一定义式：

$$T=-\varepsilon\Big/\left(k\ln\frac{n}{n_0}\right) \tag{6-2-2}$$

在一般热力学系统中，具有高能量的粒子数 $n$ 是小于具有低能量的粒子数 $n_0$ 的，即 $n<n_0$，所以 $\ln\frac{n}{n_0}<0$，代入式(6-2-2)可知，系统的绝对温度为正，即 $T>0$。但是，在特殊条件下确实也存在另一种可能，如果能使具有高能量的粒子数 $n$ 大于低能量的粒子数 $n_0$，即发生粒子数的反转 $n>n_0$，则 $\ln\frac{n}{n_0}>0$，代入式(6-2-2)可知，这时系统的绝对温度 $T<0$，即系统具有负绝对温度。由此可见，量子统计关于温度的微观定义中引入负绝对温度是合理的。

**3. 热力学第一定律的绝对温度定义与负绝对温度的含义**

热力学第一定律的基本微分方程为 $T\mathrm{d}S=\mathrm{d}U+p\mathrm{d}V$，式中 $S$ 是系统的熵(entropy)，$U$ 是系统的内能，$p$ 是系统的压强，$V$ 是系统的体积。当参量 $V$ 保持不变时，可得绝对温度的另一宏观定义式，即：

$$T=1\Big/\left(\frac{\partial S}{\partial U}\right)_V \tag{6-2-3}$$

对于孤立系统，在一般情况下，熵是随内能增加而增加的，所以 $\left(\frac{\partial S}{\partial U}\right)_V>0$，由式(6-2-3)可得 $T>0$，即系统具有正绝对温度。但是，在特殊条件下，如果系统的熵 $S$ 是系统能量 $U$ 的单调减函数，即 $S$ 随 $U$ 的变化率取负值时，即负熵系统(也就是所谓的耗散结构系统)，则有 $\left(\frac{\partial S}{\partial U}\right)_V<0$，由式(6-2-3)可得 $T<0$，这时系统具有负绝对温度。

因此，根据玻耳兹曼的能量分布律(微观理论)和热力学定律(宏观理论)，我们可以把温度的定义延拓到负绝对温度状态。

**4. 负绝对温度概念的讨论**

下面通过对自旋为$\frac{1}{2}$的理想自旋粒子系统和产生激光的粒子数反转系统的分析，进一

步说明负绝对温度的概念。

1) 对自旋为$\frac{1}{2}$的理想自旋粒子系统的分析

设有 $N$ 个自旋为$\frac{1}{2}$、固有磁矩为 $\mu$ 的粒子组成的理想自旋系统,粒子位于晶格点阵的 $N$ 个格点上。在外磁场 $\boldsymbol{B}$ 的作用下,每个粒子的磁矩有两种可能取向:平行于外场(能量为$-\mu\boldsymbol{B}$)和反平行于外场(能量为$+\mu\boldsymbol{B}$)。当有 $n$ 个粒子的磁矩平行于外场时,系统的能量为

$$E(n)=(N-n)\mu\boldsymbol{B}-n\mu\boldsymbol{B}=(N-2n)\mu\boldsymbol{B} \tag{6-2-4}$$

磁矩平行于外场的 $n$ 个粒子在 $N$ 个格点上有许多不同的排列方法,每一种不同的排列方法对应于具有相同能量 $E(n)$的不同微观态 $W$。利用排列组合知识容易证明 $n$ 个粒子在 $N$ 个格点上作不同的排列方法的总数,即与 $E(n)$对应的微观态数,也就是热力学概率

$$W=\frac{N!}{n!\ (N-n)!} \tag{6-2-5}$$

代入熵的玻耳兹曼公式 $S=k\ln W$,得 $S(n)=k\ln\frac{N!}{n!\ (N-n)!}$,考虑斯特林近似公式 $\ln n!=n\ln n-n$,$(n>10)$,得熵 $S$ 对 $n$ 的依赖关系

$$S(n)=k[N\ln N-n\ln n-(N-n)\ln(N-n)] \tag{6-2-6}$$

从而得到

$$\frac{1}{T}=\frac{\partial S}{\partial U}=\frac{\partial S}{\partial n}\times\frac{\partial n}{\partial U}=\frac{\partial S}{\partial n}\times\frac{\partial n}{\partial E}=k\ln\frac{N-n}{n}\times\left(-\frac{1}{2\mu\boldsymbol{B}}\right) \tag{6-2-7}$$

考虑式(6-2-4)、(6-2-5)、(6-2-6)和式(6-2-7)可知:

当 $n=\frac{N}{2}$时,即低能级粒子数等于高能级粒子数,即粒子数随能级作平均分布时,有 $E=0$、$W=2^N$、$S=Nk\ln 2$ 和 $T=\pm\infty$K。这表明,当各有一半粒子的磁矩平行和反平行于外场时,对应的微观数最多,系统最无序,熵有极大值,温度为$\pm\infty$K,如图 6-2-1 所示。

当 $n=N$,即高能级粒子数为零,有 $E=-N\mu\boldsymbol{B}=-N\varepsilon_1$、$W=1$、$S=0$ 和 $T=+0$K。即 $N$ 个粒子的磁矩全部平行于外场时,系统只有一个微观态,系统完全有序,熵为零,温度为$\pm 0$K,如图 6-2-2 所示。

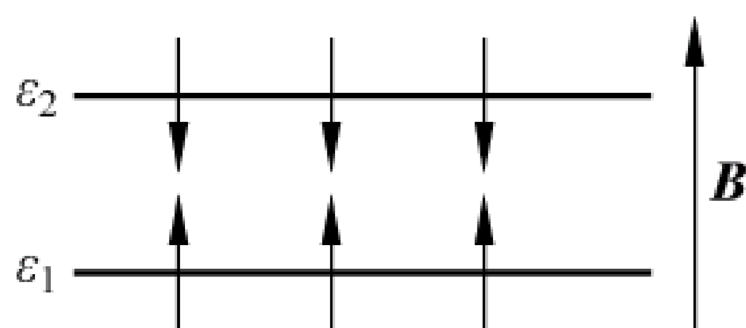



图 6-2-1 系统粒子数分布:$n_1=n_2=n=N/2$,$T=+\infty$K 时

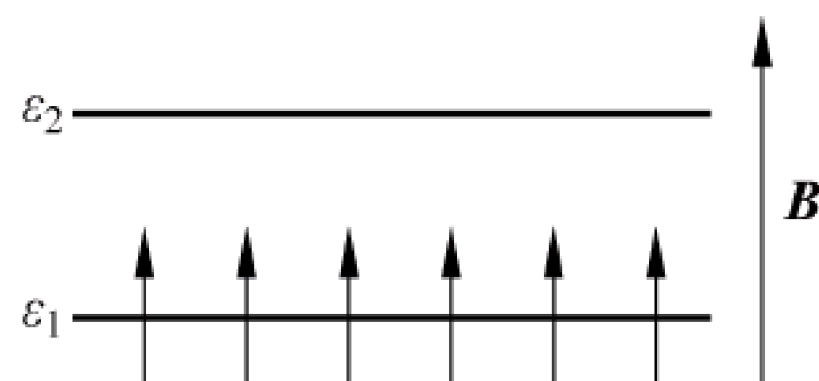



图 6-2-2 $T=+0$K 时,$n=n_1=N$,$n_2=0$;$U_{\min}=N\varepsilon_1$;$S=0$

当 $n=0$,即高能级粒子数为零,$E=+N\mu\boldsymbol{B}=+N\varepsilon_2$、$W=1$、$S=0$ 和 $T=-0$K。即 $N$ 个粒子的磁矩全部反平行于外场时,系统也只有一个微观态,系统完全有序,熵为零,温度为$\pm 0$K,如图 6-2-3 所示。

当 $n>\frac{N}{2}$时,等式(6-2-7)为正,$T>0$K,即低能级粒子数大于高能级粒子数,即粒子数

的正常分布；当 $n<\frac{N}{2}$时，等式右方为负，$T<0\text{K}$，这与正温度下粒子按能量的分布情况相反，即是粒子数反转分布，如图 6-2-4 所示。

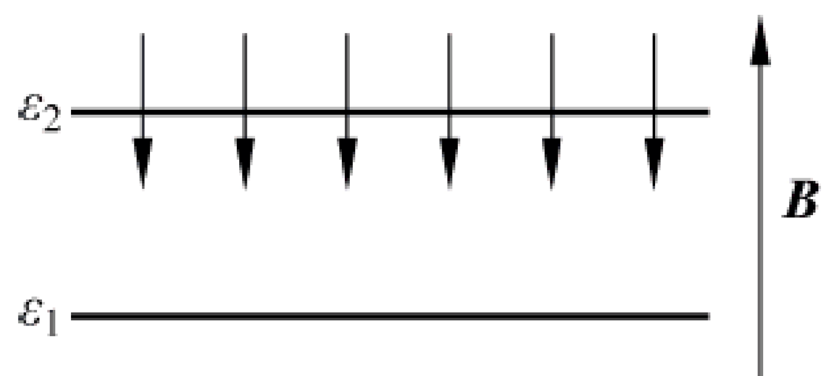



图 6-2-3 系统 $T=-0\text{K}$ 时，$n=n_1=0$，$n_2=N$；$U_{\max}=N\varepsilon_2$；$S=0$

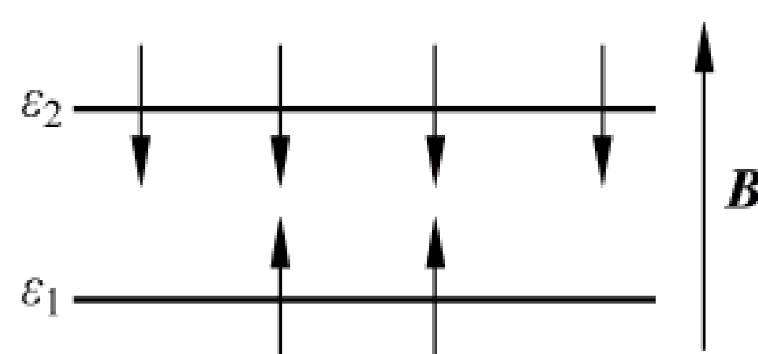



图 6-2-4 $T<0\text{K}$ 时，$n_2>n_1=N/2$；$U\uparrow$；$S\downarrow$

2）激光产生系统粒子数反转过程的分析

增益介质的粒子数反转状态是激光产生的必要条件，因此激光器系统中的工作物质的粒子数反转现象，也存在负绝对温度概念。

考虑给定物质（增益介质如红宝石）的任意两个能级 $E_1$ 和 $E_2$，$n_1$ 和 $n_2$ 分别表示两能级上的粒子数，在热平衡时，能级的粒子数由玻耳兹曼统计分布所描述，$\frac{n_1}{n_2}=\mathrm{e}^{-(E_1-E_2)/kT}$。

通常情况下 $T>0$，所以 $n_2<n_1$。但是，激光器增益介质的粒子能级在泵浦源的作用下，能够获得 $E_2>E_1$ 的状态，即粒子数反转状态：$n_2>n_1$。此时会产生受激辐射，即增益介质对电磁辐射将起放大作用，这正是粒子数反转时的负温状态，即 $T<0$。这个处于负温状态的放大过程是激光产生的基本原理。

3）负绝对温度的量子统计定义

由以上两个例子的分析可知，温度或者其热力学温度是表征粒子数按能级分布情况的物理量，而负绝对温度对应的是粒子数的反转分布（粒子数反转）状态，即当系统中处于高能级的粒子数大于处于低能级的粒子数时，就出现了负绝对温度现象。

我们考虑前面的自旋为 1/2 的理想自旋粒子系统，忽略自旋之间的相互作用，则由量子力学可知系统的哈密顿量为 $H=E=\sum_i I_i\varepsilon$，式中 $I_i$ 为第 $i$ 个自旋的自旋量子数，$I_i=\pm\frac{1}{2}$；$\varepsilon=-\mu_B B$，其中 $\mu_B$ 为玻尔磁子，$B$ 为加在系统（二级能级）上的恒定磁场。我们容易算出粒子数按能级分布的配分函数

$$Z=\sum_{(I_i)}\exp\left(-\beta\sum_i I_i\varepsilon\right) \tag{6-2-8}$$

因为 $Z=\sum_{(I_i)}\prod_{(i=1)}^{N}\exp\left(-\beta\sum_i I_i\varepsilon\right)=\prod_{(i=1)I}^{N}\sum\exp\left(-\beta\sum_i I_i\varepsilon\right)=z^N$，其中 $z=\sum_{I_i}\exp\left(-\beta\sum_i I_i\varepsilon\right)=\mathrm{e}^{-\frac{1}{2}\beta\varepsilon}+\mathrm{e}^{\frac{1}{2}\beta\varepsilon}$，则有 $\ln Z=N\ln z=N\ln(\mathrm{e}^{-\frac{1}{2}\beta\varepsilon}+\mathrm{e}^{\frac{1}{2}\beta\varepsilon})$，于是得出

$$\bar{E}=-\frac{\partial}{\partial\beta}\ln Z=N\left(\frac{1}{2}\varepsilon\right)\mathrm{th}\left(-\frac{1}{2}\beta\varepsilon\right) \tag{6-2-9}$$

$$S=k\ln Z+\frac{\bar{E}}{T}=Nk\ln(\mathrm{e}^{-\frac{1}{2}\beta\varepsilon}+\mathrm{e}^{\frac{1}{2}\beta\varepsilon})+Nk\left(\frac{1}{2}\beta\varepsilon\right)\mathrm{tch}\left(-\frac{1}{2}\beta\varepsilon\right) \tag{6-2-10}$$

从式（6-2-9）和式（6-2-10）可以得出

$$\frac{1}{T}=\frac{\partial S}{\partial U}=\frac{\partial S}{\partial E}\begin{cases}>0, & U<0\\ =0, & U=0\\ <0, & U>0\end{cases} \tag{6-2-11}$$

即为存在负绝对温度的典型热力学系统。

4）负绝对温度的实现条件

拉姆赛曾指出，负绝对温度的实现应满足的条件是：

其一，热平衡态条件。热平衡，指系统间仅在热接触条件下(即未做功)达到的共同的状态，在这种状态下，系统的宏观性质不随时间变化。

无论正绝对温度态或负绝对温度态(以下分别简称正温态和负温态)，都必须是热平衡态。因为温度是热平衡态参量，离开热平衡态，正、负绝对温度都将失去意义。

其二，能量上限条件。系统能级数目有限，且能量有一上限。能量上限条件可由$\left(\frac{\partial S}{\partial U}\right)<0$得出。如果系统能量没有上限，即存在无限多能态可供粒子占据，则随着内能增加，粒子在各能态上分布必将趋于更均匀，系统熵不可能减小，不会达到负温态。只有当系统能量存在上限，能态数目为有限时，随着内能增大、粒子在有限能态上的分布可能趋于不均匀，有序性增大，熵减小，从而可能实现负绝对温度态。

必须注意，任一个独立的物理系统，其总能量原则上没有上限(例如，粒子动能一般没有上限)，因此，不能对一个独立的物理系统整体引入负温度。但系统的某一自由度能量可能有上限，故可对这一自由度引入负的绝对温度。

其三，与正绝对温度系统隔绝。设系统 1 与系统 2 组成一个孤立系统，$\delta E_{1-2}$ 为系统 1 流入系统 2 的能量，两个系统熵的和为 $S=S_1+S_2$，由热力学第二定律得

$$\delta S=\left(\frac{\partial S_2}{\partial E_2}-\frac{\partial S_1}{\partial E_1}\right)\delta E_{1-2}=k(\beta_2-\beta_1)\delta E_{1-2}>0 \tag{6-2-12}$$

当 $\beta_1<\beta_2$，$T_1$ 为负绝对温度，即系统 1 处于负温态，由式(6-2-12)得 $\delta E_{1-2}>0$，即能量由系统 1 流向系统 2，也就是能量由负绝对温度系统(如核磁矩系统)流向正绝对温度系统(如晶格振动系统)。这种流向是由于孤立系统的熵 $S$ 总量向增加的方向发展的原因。所以，负温态系统一旦与正温态系统接触，便会由于能量交换而返回正温态，因此负温态系统应当是隔绝的。这里所谓“隔绝”，应理解为相互作用很弱。

**5. 负绝对温度的特点**

通常我们认为，“热”的物体其温度具有更大的代数值，当我们分别考虑正、负两种温度时，这种定义显然是正确的。一旦把二者联系起来，就必然引起混乱，从正绝对温度过渡到负绝对温度，系统能量单调增加，负温区不是在正的“绝对零度”以下，而是位于“无限大温度”以上。负温区的热激发能力比正温区更大，一个负温系统和一个正温系统热接触，热量将从负温系统传向正温系统。温标从冷到热排列如下：

$T=+0\text{K}\rightarrow T>0\text{K}\rightarrow T=\pm\infty\text{K}\rightarrow T<0\text{K}\rightarrow T=-0\text{K}$，如图 6-2-5 所示。

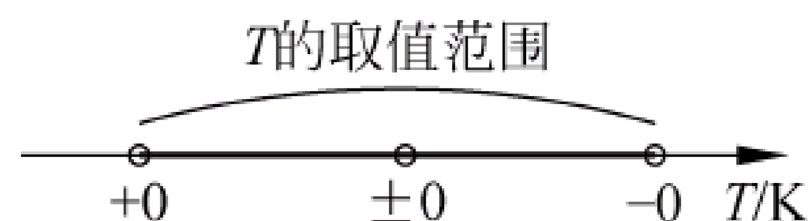



图 6-2-5 $+0\text{K}\rightarrow T=\pm\infty\text{K}\rightarrow T=-0\text{K}$

特点 1：$T=+\infty\text{K}$ 实际为热力学温度正、负温区的交汇点；

特点 2：$T=+\infty\text{K}$ 与 $T=-\infty\text{K}$ 实际是同一热学状态，处于同一温度；

特点 3：负绝对温度($T<0$)比正绝对温度($T>0$)高，而且比$\pm\infty$K 还高，$T=\pm 0$K 仅是中等温度；

特点 4：由正温状态变为负温状态，不是将系统“冷却”到负温，而是“加热”到负温；

特点 5：温度$+0$K 和$-0$K 是两个完全有序态，差异很大。在孤立系统(不与外界交换能量)中，$+0$K 状态可保持任意长时间，是一种稳定平衡态，在此状态下，系统能量为最小值，但该状态仅靠系统本身是无法实现的。$-0$K 状态则是极不稳定状态，因为此时系统能量为最大值，若无外界作用，系统将立即改变这种极不稳定态，高能级粒子将自发地降到各低能级上，也就是说，负绝对零度状态是个极不稳定的平衡态。

**6. 负绝对温度概念对热学概念及规律的扩展**

负绝对温度中的温度是广义的温度概念，热力学的有关规律都是在正绝对温度系统中建立的，但在负绝对温度系统中也多数成立，只是某些地方必须加以修正。

1）温度概念的修正

（1）温度的宏观定义的修正

对负绝对温度系统，温度的宏观定性定义仍表述为温度是系统达到热平衡时的标志，它是决定一物质系统是否与其他系统处于热平衡的宏观性质。其特征就在于：一切互为热平衡的系统都具有相同的温度。

宏观定量定义的修正：关于温度的宏观定量定义可以不做改变。因为负绝对温度状态是在原热力学温度超越$+\infty$K，继续向更高温度区的延伸，所以，热力学温标的测温依据 $Q$ 和标度法(以水的三相点为固定点，并规定了一度的大小)均可不变。即正、负温区是统一的热力学温标。但热力学温标的原定义式$\dfrac{Q}{Q_{tr}}=\dfrac{T}{273.16\text{K}}$只能在正温区使用，当在正温区由此确定热力学温标后，再超越$+\infty$K 温度对负温区刻度，从而构成完整的热力学温标。

（2）温度微观定义的修正

温度的经典微观定量定义 $\bar{\omega}=\dfrac{1}{2}m\overline{v^2}=\dfrac{3}{2}kT$，否定了 $T<0$ 的存在，应以量子统计的结论替代，即 $n=n_0\mathrm{e}^{-\frac{\varepsilon}{kT}}$。相应地，温度的微观定性定义可以较明确地表达为：温度是表征粒子数按能级分布情况的物理量。

2）热力学定律表述的修正

（1）热力学第一定律的表述

因为热力学第一定律是能量守恒和转化定律，是自然界的普遍定律，在负绝对温度范围内也适用。在热力学第一定律中，热量、内能等概念在正、负绝对温度系统中都不改变。只是由于系统中的能量在负绝对温度系统中比正绝对温度系统中还要大，因而从正绝对温度系统过渡到负绝对温度系统需要外界提供能量。无论在哪一种系统中，能量都不会凭空产生，只能从一种形式转化为另一种形式，而且转化过程中总能量保持不变。

（2）热力学第二定律的克劳修斯表述的修正

热力学第二定律的克劳修斯表述不变：不可能把热量从低温物体传到高温物体而不引起其他影响。设想在负绝对温度系统中存在一高温热源 $T_1=-100$K，低温热源 $T_2=-200$K，此过程中，热量从低温度热源 $T_2$ 传到高温热源 $T_1$，则其熵变：

$$\Delta S=\frac{Q}{T_1}-\frac{Q}{T_2}=Q\left(\frac{1}{T_1}-\frac{1}{T_2}\right)=Q\left(\frac{1}{-100}-\frac{1}{-200}\right)=Q\times\frac{1}{-200}<0$$

这个过程 $\Delta S$ 小于 0,即不符合熵增加原理,也就是说这个过程不可能自发实现。因此热力学第二定律的克劳修斯表述在负温度系统中是成立的。

(3) 热力学第二定律的开尔文表述的修正

热力学第二定律的开尔文表述需要修正。

可以设想卡诺循环过程来说明这个问题。假设有两个负温热源:$T_1=-100\mathrm{K}$ 和 $T_2=-200\mathrm{K}$,工作于其间的一卡诺循环热机从高温热源 $T_1$ 吸收热量 $Q_1$,在低温热源 $T_2$ 放热 $Q_2$,对外做功,则其效率为 $\eta=1-\dfrac{T_2}{T_1}=1-\dfrac{-200}{-100}=-1<0$。

这意味着热机从高温热源吸热后不是对外界做功,而是需要外界对热机做功;而在逆循环中,$\varepsilon=1-\dfrac{T_1}{T_2}=1-\dfrac{-100}{-200}=\dfrac{1}{2}>0$。这表示制冷机对外做功。于是我们可以设想这样一个循环过程:如图 6-2-6 所示,热机从低温热源吸收热量 $Q_2$,在高温热源放热 $Q_1$,同时对外做功 $W=Q_2-Q_1$,然后 $Q_1$ 又流回低温热源,结果热机从单一热源吸热完全转变为功而不产生其他影响成为可能,而把功转变为热倒是不可能的,开尔文说法不再成立。可以修正为:不可能制造一种机器,它能从一个正绝对温度热源吸取热量使之完全转变为功;或者通过外界对它做功把热量传递给一个负绝对温度热源而不产生其他影响。

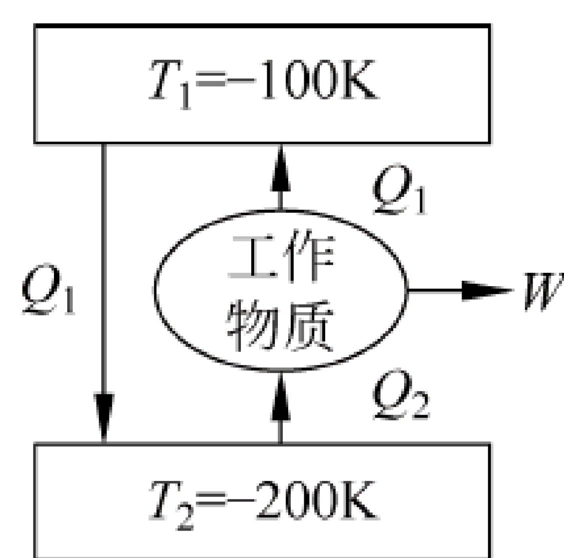



图 6-2-6 热力学第二定律的开尔文表述的修正

正是由于负绝对温度与正绝对温度概念上的差别,致使在负绝对温度区开氏说法与正温区恰好相反。

(4) 热力学第三定律(负绝对温度)的表述

热力学第三定律(不可能使一个物体冷到绝对温度的零度)也没有改变,不论是 $T=+0$ 还是 $T=-0$,都需要消耗无限多的能量,都是达不到的,这一定律在正、负绝对温度系统中都是适用的。而温度 $T=\pm\infty\mathrm{K}$ 却是可以实现的。

**7. 结论**

负绝对温度概念是热力学绝对温度概念的推广概念,是热力学系统的量子描述的产物,它是经典物理学描述所不能给出的。负绝对温度概念的物理本质与系统的实际“冷热”没有直接关系,而是与系统的能量和熵的变化有直接关系。

负绝对温度概念的提出,将使我们对温度概念的认识又向前推进了一步。

从理论上来看,绝对温度的负温区的特点已经在核自旋、顺磁电介质、量子放大器、激光器和核磁共振等研究中取得了一些成就。用负绝对温度的相对性及熵概念来研究宇宙的起源、白矮星和黑洞等物理现象也是目前人们感兴趣的研究领域。

## 6.3 不可逆过程和熵

热力学的不可逆过程,可由热力学第二定律给出在不同的外界条件下,系统内部不可逆过程进行方向的判别标准。如果不可逆过程的初态和终态是平衡态,则可根据热力学关系求过程的总效应。平衡态热力学不能讨论过程进行的速率,而不可逆过程热力学能够讨论过程进行的速率问题。在客观世界中,非平衡的热力学系统(系统可以是物理的、化学的、生物学的和社会学的)与不可逆的热力学过程是大量存在的,非平衡态热力学在热传导、扩散、化学变化、天体演化、生物的物理和化学过程直至社会经济的变化过程的研究中都有着非常

重要的作用。因此,对热力学不可逆过程的研究是非常重要的。本节对热力学系统内进行的不可逆过程——热传导和扩散等输运过程的性质进行了分析,并对系统的熵的变化从熵流密度和局域熵产生率两个方面进行了比较详细的讨论,给出了具体的不可逆过程求熵流密度和局域熵产生率的表达式,这对理解热力学不可逆过程的物理意义是很有帮助的。

### 6.3.1 非平衡系统的局域平衡假设

我们已经知道,热力学平衡态是由一组宏观参量来描述的,同样非平衡态也可以用一组宏观参量来描述,但这些宏观参量与空间位置及时间有关,如浓度、密度等。对于热力学平衡态,已经证明存在三个基本的热力学函数即温度、内能与熵,并由此引出焓、自由能、吉布斯函数等辅助态函数。对于非平衡态,这里只讨论近平衡区的线性不可逆过程,因此,可采用局域平衡的假定,此假定的具体定义是:把系统划分为许多小部分,$\rho_i=\rho_i(\boldsymbol{r},t)(i=1,2,\cdots,k)$每一小部分从宏观来看是很小的,从微观来看却包含足够多的粒子,每一小部分的弛豫时间与观察的时间和整个系统的弛豫时间相比非常小。整个系统的状态虽然是非平衡的,但对局域来说则可近似认为仍然处于平衡状态,热力学函数仍有意义。并且在每一时刻这些局域热力学量的变化仍然遵从可逆过程的热力学基本方程

$$T\mathrm{d}S=\mathrm{d}U+p\mathrm{d}V-\mathrm{d}W'-\sum_i u_i\mathrm{d}n_i \tag{6-3-1}$$

式中,$u_i$ 是第 $i$ 种组元的化学势,$\mathrm{d}W'$是外界对所研究部分所做的非膨胀功。对于强度量,如温度 $T$,每一局域仍可定义温度。但就整个系统来说并不存在一个统一的温度,而是存在一个按空间与时间的连续分布,即形成一个温度场 $T(\boldsymbol{r},t)$。类似地可以定义其他的强度量。对于可加量熵,在每一局域既然仍有定义,那么整个系统的熵为各局域熵的总和,若空间某体积元的局域熵表为

$$S_V=S_V(\rho_1(\boldsymbol{r},t),\rho_2(\boldsymbol{r},t),\cdots) \tag{6-3-2}$$

则整个系统处于非平衡态时的熵为

$$S=\int S_V\mathrm{d}V \tag{6-3-3}$$

类似地,亦可定义非平衡态时的其他可加热力学量,如内能 $U=\int U_V\mathrm{d}V$、自由能 $F=\int F_V\mathrm{d}V$ 和吉布斯函数 $G=\int G_V\mathrm{d}V$。

### 6.3.2 热流和熵产生率

对于单纯的热传导现象,设一个孤立系统仅仅由两个固定体积的小部分组成,该两个小部分有不同的温度,它们之间用一导热隔板隔开,现在这两个小部分只有热量的传递。整个系统的熵变等于两个部分的熵变之和,并且有 $U_1+U_2=C$,所以

$$\mathrm{d}S=\mathrm{d}S_1+\mathrm{d}S_2=\frac{\mathrm{d}U_1}{T_1}+\frac{\mathrm{d}U_2}{T_2}=\left[\frac{1}{T_1}-\frac{1}{T_2}\right]\mathrm{d}U_1 \tag{6-3-4}$$

或熵增加率即为$\dfrac{\mathrm{d}S}{\mathrm{d}t}=\left[\dfrac{1}{T_1}-\dfrac{1}{T_2}\right]\dfrac{\mathrm{d}U_1}{\mathrm{d}t}$,即

$$S^0=\Delta\left[\frac{1}{T}\right]U_1^0=\Delta\left[\frac{1}{T}\right]Q^0 \tag{6-3-5}$$

其中,$\Delta\left[\dfrac{1}{T}\right]=\Delta\left[\dfrac{1}{T_1}-\dfrac{1}{T_2}\right]$,$Q^0$ 是单位时间内从高温部分流向低温部分的热量,称为热流。

这里只有热量传输，所以，每个小部分的内能增加率等于进入该部分的热流。而直接导致热流的原因是存在温度差或$\frac{1}{T}$的不均匀性，即热流是温度不均匀性的结果。上式表明系统的熵增加率取决于两个因素——原因 $\Delta\left[\frac{1}{T}\right]$ 和结果 $Q^0$ 的乘积。

如果除了温度不均匀以外，物质的性质(例如化学性质和电学性质)也不均匀，则除了热传导外，还同时有物质的迁移。设两部分的温度和化学势分别为 $T_1, \mu_1$；$T_2, \mu_2$。注意到能量守恒 $U_1+U_2=C_1$ 和粒子数守恒 $N_1+N_2=C_2$，得到系统的熵增加率为

$$S^0=S_1^0+S_2^0=\frac{\partial S_1}{\partial U_1}U_1^0+\frac{\partial S_2}{\partial U_2}U_2^0+\frac{\partial S_1}{\partial N_1}N_1^0+\frac{\partial S_2}{\partial N_2}N_2^0$$

$$=\left[\frac{1}{T_1}-\frac{1}{T_2}\right]U_1^0+\left[\frac{\mu_2}{T_2}-\frac{\mu_1}{T_1}\right]N_1^0=\Delta\left[\frac{1}{T}\right]U^0-\Delta\left[\frac{u}{T}\right]N^0$$

式中已令 $U_1^0\equiv U^0$、$N_1^0\equiv N^0$，$N^0$ 是单位时间内从化学势高(或压强高)的部分流向化学势低(或压强低)的部分的粒子数，称为粒子流；$U^0$ 是单位时间内从温度高的部分流向温度低的部分的能量，称为能流。表明两个小部分之间同时有温度差和化学势差的情形下，将同时发生热流和粒子流，而熵增加率是各个“原因”和对应“结果”乘积的和。但这时的能流 $U^0$ 已经和单独只有温度差时的不同。因为对于任何一个小部分，能流 $U^0$ 来源于两个方面，一方面因为有温度差发生热传导现象，故有热流 $Q^0$ 流入(或流出)所考虑的这个小部分；另一方面，因为有化学势差，故有粒子流 $N^0$，而伴随粒子流就有能流 $\mu N^0$ 流入(或流出)所考虑的这个小部分。所以，一般有 $U^0=Q^0+\mu N^0$，上述结果可以推广到由许多小部分组成的系统，即讨论物性在空间连续变化的情形。如果系统还处在外电场中，设某点的电势为 $\psi$，当某一点的带电粒子数密度改变 $\mathrm{d}n$ 时，外界对系统所做的功是 $\mathrm{d}W=e\psi\mathrm{d}n$ 或$\frac{\mathrm{d}W}{\mathrm{d}t}=e\psi\frac{\mathrm{d}n}{\mathrm{d}t}$，这时熵密度的增加率为(一种组元的系统)：

$$\frac{\mathrm{d}S}{\mathrm{d}t}=\frac{1}{T}\frac{\partial u}{\partial t}-\frac{u}{T}\frac{\partial n}{\partial t}-\frac{1}{T}\frac{\mathrm{d}W}{\mathrm{d}t}\tag{6-3-6}$$

由连续性方程：$\frac{\partial\mu}{\partial t}=-\nabla\cdot\boldsymbol{J}_v$($\mu$ 是内能密度，$\boldsymbol{J}_v$ 是能流密度)和$\frac{\partial n}{\partial t}=-\nabla\cdot\boldsymbol{J}_n$($n$ 是粒子数密度，$\boldsymbol{J}_n$ 是粒子流密度)，得

$$\frac{\partial S}{\partial t}=\frac{1}{T}\frac{\partial\mu}{\partial t}-\frac{\mu+e\psi}{T}\left[\frac{\partial n}{\partial t}\right]=-\frac{1}{T}\nabla\cdot\boldsymbol{J}_v+\frac{\mu_e}{T}\nabla\cdot\boldsymbol{J}_n\tag{6-3-7}$$

其中 $\mu_e=\mu+e\psi$ 为电化学势，其物理意义是空间某个小部分中每增加一个带电粒子所伴随增加的能量。

由式 $T\mathrm{d}S=\mathrm{d}u+p\mathrm{d}V-\sum\limits_i\mu_i\mathrm{d}n_i$ 可知，当粒子数增加 $\mathrm{d}n$ 时，内能增加 $\mu\mathrm{d}n$，因此当存在带电粒子流时，内能流密度 $\boldsymbol{J}_\mu$ 可表示为 $\boldsymbol{J}=\boldsymbol{J}_q+\mu\boldsymbol{J}_n$。利用公式：$\nabla\cdot(A\boldsymbol{B})=A\ \nabla\cdot\boldsymbol{B}+\boldsymbol{B}\cdot\nabla A$，可知$\frac{\partial S}{\partial t}=-\nabla\cdot\left[\frac{\boldsymbol{J}_m}{T}-\frac{m_e}{T}\boldsymbol{J}_m\right]+\boldsymbol{J}_n\times\tilde{N}\left[\frac{1}{T}\right]-\boldsymbol{J}_n\times\tilde{N}\left[\frac{m_e}{T}\right]$；再令熵流密度矢量为：$\boldsymbol{J}_S=\frac{1}{T}\boldsymbol{J}_m-\frac{m_e}{T}\boldsymbol{J}_n=\frac{1}{T}\boldsymbol{J}_q$ 以及令局部熵密度产生率为 $Q=\boldsymbol{J}_m\times\tilde{N}\left[\frac{1}{T}\right]-\boldsymbol{J}_n\times\tilde{N}\left[\frac{m_e}{T}\right]=\boldsymbol{J}_q\times\tilde{N}\left[\frac{1}{T}\right]-\boldsymbol{J}_n\times\frac{1}{T}\tilde{N}m_e$，则有

$$\frac{\partial S}{\partial t}=-\tilde{N}\times \boldsymbol{J}_S+Q=-\tilde{N}\times \frac{\boldsymbol{J}_q}{T}+\boldsymbol{J}_q+\boldsymbol{J}_q\times \tilde{N}\left[\frac{1}{T}\right]-\boldsymbol{J}_n\times \frac{1}{T}\tilde{N}m_e \tag{6-3-8}$$

### 6.3.3　结论

上面的讨论表明，熵密度增加率可分为两部分：$-\nabla\cdot \boldsymbol{J}_S$ 是从体积元外流入的热量所引起的局部熵密度的增加率；$Q$ 是体积元中的热传导过程和物质输运过程所引起的局部熵密度产生率。因此整个非平衡系统熵的变化是由两部分构成，即 $\mathrm{d}S=\mathrm{d}S_e+\mathrm{d}S_i$，其中 $\mathrm{d}S_e$ 表示系统与环境交换物质和能量所引起的系统熵的变化，称为熵流。根据热力学第二定律可知 $\mathrm{d}S\geqslant \mathrm{d}S_e$，即不可逆过程中非平衡总熵的变化大于熵流 $\mathrm{d}S_e$，这是因为非平衡系统的内部存在着由不可逆过程(如电流、热传导、热扩散等)引起的系统内部的熵产生 $\mathrm{d}S_i$，按照热力学第二定律，应有 $\mathrm{d}S_i\geqslant 0$，其中对于不可逆过程取不等号；对于可逆过程取等号，表示系统内部的熵产生总是正的。对于孤立系统 $\mathrm{d}S_e=0$，故得 $\mathrm{d}S=\mathrm{d}S_i\geqslant 0$，这就是熵增加原理。对于开放系统，$\mathrm{d}S_i$ 的符号永远为正，但 $\mathrm{d}S_e$ 的符号则是不定的，当 $\mathrm{d}S_e=-\mathrm{d}S_i$ 时，$\mathrm{d}S=0$，这表示系统的熵处于极值，这时开放系统可维持在一个非平衡的定态。

## 6.4　热熵与信息熵

现代统计力学的发展已超越了物理科学的范围，渗透到化学、生物学、社会学等各个科学领域，并引起人们对宇宙学和哲学的根本问题产生新的认识和讨论。比如，信息论中的熵与热力学中的熵——将二者分别称为信息熵和热熵，尽管它们出现在完全不相联系的两个学科中，但都被叫作熵，是因为它们具有相同的数学性质，它们表述的是物质世界的同一属性。熵的概念同时在热力学和信息论这样毫不相关的学科领域中出现，看来绝非偶然。因为它是一个研究大量事件总体规律的量。单个物体的运动遵循动力学规律，大量单个物体的共同运动在其总体上体现出特有的规律性。在这方面，有序和无序也许是最基本最粗略的描述，熵则是对这种无序性给予度量。

### 6.4.1　热熵

热力学中的态函数熵，是克劳修斯于 1865 年定义的。对于一个热力学系统，其循环过程的克劳修斯等式：

$$\oint \frac{\partial Q}{T}=0 \tag{6-4-1}$$

如图 6-4-1 所示循环过程中，任意选定两个平衡态 A 和 B，而将整个循环过程分为Ⅰ、Ⅱ两部分，则有：$\oint \frac{\partial Q}{T}=\int_{\mathrm{I\,A}}^{\mathrm{B}} \frac{\partial Q}{T}+\int_{\mathrm{II\,B}}^{\mathrm{A}} \frac{\partial Q}{T}=0$。

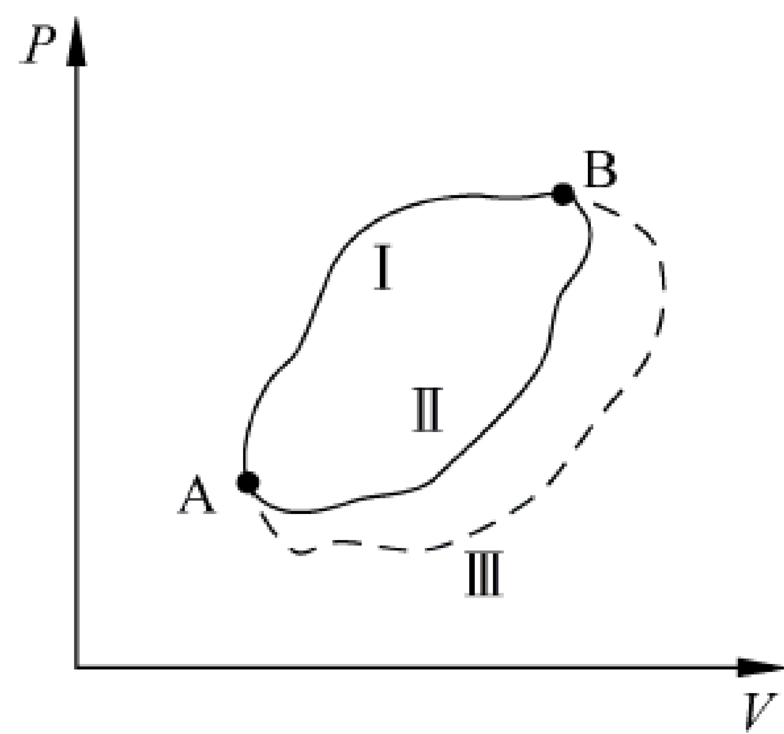



图 6-4-1　循环过程

由 $\int_{\mathrm{I\,A}}^{\mathrm{B}} \frac{\partial Q}{T}-\int_{\mathrm{II\,A}}^{\mathrm{B}} \frac{\partial Q}{T}=0$ 得

$$\int_{\mathrm{I\,A}}^{\mathrm{B}} \frac{\partial Q}{T}=\int_{\mathrm{II\,A}}^{\mathrm{B}} \frac{\partial Q}{T} \tag{6-4-2}$$

同理，可以证明 A、B 间任意准静态过程Ⅲ、Ⅳ、Ⅴ……，有

$$\int_{\mathrm{I\,A}}^{\mathrm{B}} \frac{\partial Q}{T}=\int_{\mathrm{II\,A}}^{\mathrm{B}} \frac{\partial Q}{T}=\int_{\mathrm{III\,A}}^{\mathrm{B}} \frac{\partial Q}{T}=\int_{\mathrm{IV\,A}}^{\mathrm{B}} \frac{\partial Q}{T}=\int_{\mathrm{V\,A}}^{\mathrm{B}} \frac{\partial Q}{T} \tag{6-4-3}$$

即积分的值 $\int_{A}^{B}\frac{\partial Q}{T}$ 与路径无关，只与初态和终态的状态有关，于是引入态函数熵 $S$，有

$$S_B - S_A = \int_{A}^{B}\frac{\partial Q}{T} \tag{6-4-4}$$

式(6-4-4)只能计算 A,B 两种平衡态之间的熵差，不能计算二态熵 $S_A$ 和 $S_B$ 的绝对数值。在热力学计算中，有时可以规定一个熵值为相对零点，由此可以计算出其他平衡态相对于这一零点的熵值。熵的绝对零点由热力学第三定律规定，该定律的能斯特表述为：当温度为绝对零度时，任何物质的熵都相等。

设对 1mol 某物质从绝对零度开始加热，在压强不变的条件下温度升高 $dT$，所需热量为：$dQ = C_T dT$，其熵的增加量为 $dS = \frac{dQ}{T} = \frac{C_T dT}{T}$，可以求得该物质在温度为 $T$ 时的熵的绝对数值为

$$S = \int_{0}^{T}\frac{C_T dT}{T} \tag{6-4-5}$$

然而，并不能用对式(6-4-5)的简单积分求出 $S$，因为 $C_p$ 本身与温度有关，特别是在低温条件下，所以，只有知道 $C_p$ 与温度(直到绝对零点)的关系，才能求出 $S$ 的绝对数值。由于绝对零度不可能达到，只能无限接近，由此得知，用式(6-4-5)来求取的熵的绝对值也只能是一个无限接近准确的数值。

熵作为描述系统平衡状态的参量函数，与另一函数内能相比较，二者都与系统包含的物质的量成正比，因而熵与内能一样具有广延性。但是，内能与其他形式的能量合在一起遵守能量守恒定律，那么熵是否也与什么别的量合起来守恒呢？我们应当明确：熵本身是不遵守任何形式守恒定律的。尽管在可逆过程中熵保持不变，但在不可逆过程中熵总是增加的，即热力学过程向着熵增加的方向进行。

值得注意的是，熵虽然与系统所包含的物质的量成正比，但是，如果把熵平均分配到构成物质系统的每一个分子、原子等微观粒子，这就失去了熵本身的含义。因为熵是一个宏观量。对于热力学研究的由大量分子、原子组成的热力学系统，熵是对整个系统而言。分子、原子作为微观粒子，它们本身的运动遵循动力学规律。每个微观粒子在与其他大量的微观粒子的共同运动中对系统的熵做出贡献，而对它们每一个个体，并不存在熵。

根据熵增加原理，一个孤立系统的熵永不减少，自发过程总是沿着熵增加的方向进行。在孤立系统发生变化时，态函数熵可能有几个极大值。在熵取最大的极大值时，系统处于稳定平衡。取较小的极大值时，系统处于亚稳平衡。

另一方面，热力学定律的统计意义说明，一个孤立系统的自发过程总是由概率小的状态向概率大的状态进行。由此，我们可以看出熵与概率存在着必然的联系。我们假设熵 $S$ 是概率 $\omega$ 的函数

$$S = f(\omega) \tag{6-4-6}$$

为了求出二者之间具体的函数关系，我们把所研究的系统分为任意两部分，它们各自的熵与概率的关系为：$S_1 \approx f(\omega_1)$ 和 $S_2 \approx f(\omega_2)$。

根据概率论，两个独立事件同时发生的概率等于各独立事件的概率之积：

$$\omega = \omega_1 \cdot \omega_2 \tag{6-4-7}$$

两系统的总熵等于二者熵之和：

$$S = S_1 + S_2 \tag{6-4-8}$$

由式(6-4-6)、式(6-4-7)、式(6-4-8)相联系得到

$$f(\omega_1 \cdot \omega_2) = f(\omega_1) + f(\omega_2) \tag{6-4-9}$$

显然，只有对数函数满足这一恒等式，于是可以写出：$f(\omega \cdot \omega) = f(\omega) = k\ln\omega$，即

$$S = k\ln\omega \tag{6-4-10}$$

这就是玻耳兹曼-普朗克关系式。式中 $k$ 为比例系数。我们可以从麦克斯韦-玻耳兹曼分布律推导出同样的关系式，并证明比例系数 $k$ 就是玻耳兹曼常数。

注意，式(6-4-10)中的 $\omega$ 是指热力学概率，根据统计物理学中的等概率原理，孤立系统中所有微观状态是等概率的，即相空间中，能量曲面 $E$ 和 $E+\Delta E$ 之间，相等体积的概率相等，其物理含义是：系统中每一微观粒子的状态可选定的可能性相同，没有任何一个粒子具有特殊性。但是，微观状态概率相等，而宏观状态的概率绝对不相等，宏观状态都是出现于所包含微观状态数最多的状态，热力学概率则定义为给定宏观状态相对应的微观状态数。

式(6-4-10)可看作是统计物理学中熵的定义式，它把系统的宏观量熵和微观性质联系起来。从这个式子可以看出，热力学系统所对应的微观状态数值越大，即热力学概率越大，熵值也越大。而微观状态数越大，就意味着微观粒子的杂乱无章程度就越大，即无序性越大。所以，熵的大小表示出系统无序程度。

## 6.4.2　信息熵

### 1. 信息熵的引入

信息是物理系统所具有的与熵相反的另一种属性。如上所述，熵被用来度量一个系统统计描述的不确定程度，而信息则是系统的有组织程度的度量。

现代信息论的奠基人，美国麻省理工学院的香农(Claude Elwood Shannon，1916—2001)征求了著名数学家冯·诺依曼(John von Neumann，1903—1957)的意见，借用统计力学中熵的概念提出了信息熵。香农把(物理学的)熵与信息紧密地联系在一起。

玻耳兹曼曾写道：“熵是一个系统失去了的‘信息’的度量。”信息熵与热熵公式相差一个负号，说明信息和熵是互补的，信息就是负熵，表示系统有序的程度。因此，信息熵就是平均信息量。

法国物理学家雷昂-布里渊(Léon Brillouin，1889—1969)在1956年发表的专著《科学与信息论》中深刻指出，信息熵与热力学中熵的数字表达式是一致的，信息熵是解除不定性所需的信息量，而热熵是系统混乱程度的度量。要使混乱的系统有序化就需要信息，而信息的丢失就表示系统混乱程度的增加，两者互为负值。

香农证明了一条定理：一个系统的微观状态数若为 $2^r$，则该系统的信息含量就是 $r$ 比特(bit)，比特作为信息的单位，表示在两个概率相等的事件中确定其一所需的信息量。

香农还确定了系统的信息 $I$ 与熵 $S$ 之和保持恒定的规律：

$$S + I = 常数 = S_{\max} = I_{\max} \tag{6-4-11}$$

我们知道，态函数内能与其他形式的能量共同遵守能量守恒定律，这里，确定系统的熵与信息之和恒定则可以用来与之相对应。尽管它们二者所涉及的对象和内容大相径庭。

其一，内能与其他形式的能量可以相互转化，在此，熵和信息也是可以相互转化的，这一点后面将会讲到。

其二，热力学第二定律提出了能量转化的方向性的问题，而信息与熵的相互转化也存在方向性，这个问题在后面叙述。

确定系统的信息与熵之和保持恒定的规律可以用一个简单的例子来说明：若某系统具

有 $2^3=8$ 个可能状态,则确定系统处于其中某一个状态而不是其他状态所需的信息量为 3 比特。这时,系统的不确定程度,即熵等于 0。如果仅能确定系统处于其中的 4 个状态中的某一个,而不会是其余 4 个中的某一个,则所具有信息量为 1 比特,于是,系统的不确定程度,即系统的熵为 2 比特,由此,也得到了一个量度熵的单位,与信息的单位相同,而且熵也即负信息,信息即为负熵。

对于信息来说,如果系统具有 $n$ 种等概率的结果,则该系统的最大信息含量为

$$I=k\ln n \tag{6-4-12}$$

式中,$k$ 为比例系数。另外,与熵增加原理相对应,有信息不增原理。

以上情况说明,信息与熵都是描述物理系统性质的物理量,它们反映系统所具有的特性的不同侧面。

**2. 信息熵的物理意义**

设想图 6-4-2 所示的实际封闭系统被隔板分为 A,B 两部分,A 部充有气体,B 部为真空。当抽去隔板,气体分子将进入 B 部,经过一段时间后,气体分子将均匀分布,系统处于客观平衡态。

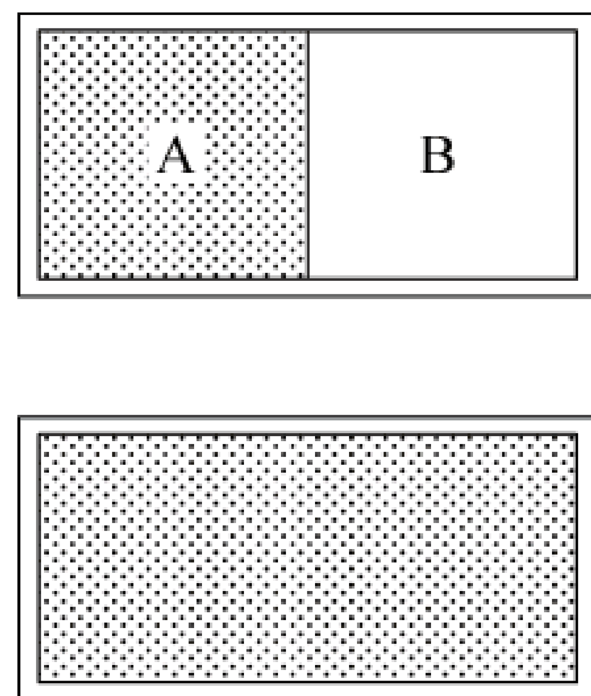


图 6-4-2 实际封闭系统

对于这一过程,我们设想在某一时刻 $t_0$,所有分子的自由运动速度方向都过来,这以后,所有分子的运动和碰撞都将决不违背动力学规律。这样,经过从抽去隔板开始到 $t_0$ 时刻相等的时间间隔。所有分子将会毫不例外地回到 A 部原来的位置上。

以上问题的关键在于,我们无法把握 $t_0$ 时刻那种特殊的初始条件,即不可能具有那么大的信息量。因为质量仅以克计的气体分子的数量级也在 $10^{20}$ 以上,这么大数量的分子的初始位置和速度,即使在测不准原理的精度范围内也是无法做到的。这就是热力学第二定律所指出的不可逆过程的充分理由。

对于一个微观粒子,采用笛卡儿坐标,任意时刻有三个空间位置坐标 $x,y,z$ 和三个速度分量 $v_x,v_y,v_z$,即我们只要知道这 6 个数据,就可以完全确定该粒子的运动状态。

对于有 $N$ 个粒子的系统,必须知道 $6N$ 个数据,这可用 $6N$ 维相空间的一个点的 $6N$ 个坐标数来表示,即该系统有 $6N$ 个自由度。

采用 $6N$ 维相空间,系统某时刻的状态对应相空间的一个点,系统的变化过程为相空间的一条曲线,这条曲线不可能有相交和分支(但可能闭合),否则在相交点和分叉点,其对应的状态将是不确定的。

为此,我们不妨将概率比喻为相空间中的流体。这种概率流体并不能完全相似于实际的流体,它在相空间的"流动"过程中必须具有以下三个特点:其一,密度不变;其二,连续不分断;其三,不分流。为表现这些特性,设想如图 6-4-3 所示概率流体在相空间的流动图,开始,如图(a),概率流体只在相空间中占据一个较小的相体积,随后,因为概率流体的流动而占据较大的相体积,如图(b),但这种流动如同拉面条一样只是变细和延长,既不存在分支,也没有稀释和浓缩,保持原有的密度,经过足够长的时间,概率流体将均匀布满限定的整个相体积,如图(c)。流体本身的总体积和密度都不变,从总体宏观上看其分布是均匀的,但从细节(微观)上看却并不均匀。

从信息的角度来看,上面表述的情况应怎样分析呢?如果整个系统应含有 $2^r$ 个微观态,其包括的总的信息量则为 $I=\log_2 2^r=r$ 比特。我们可以把终态的概率流体分布的相空间分为 $2^r$ 个微小的相体积,每个小体积中包含等量的概率流体。这样,要确定流体在整个

相空间的分布所需的信息量仍为 $r$ 比特。可见，初态中的宏观信息转变为终态中的微观信息。

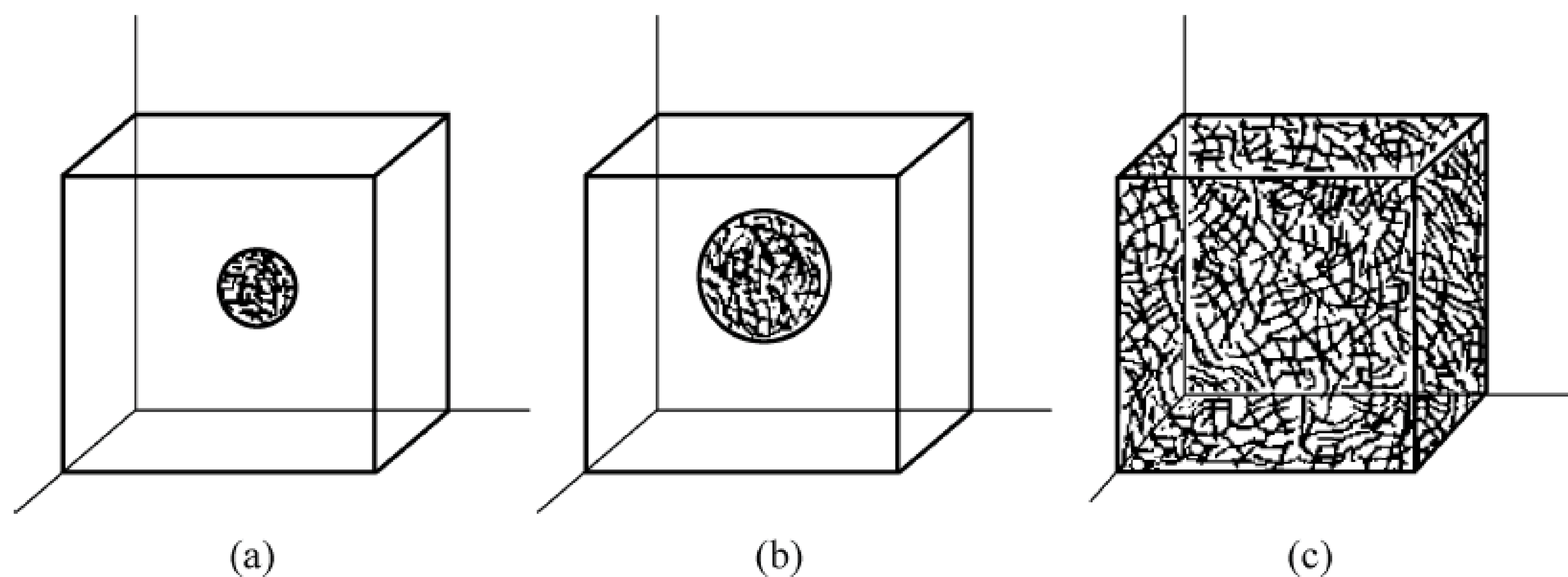



图 6-4-3 概率流体在相空间的流动

上述计算导致微观信息随宏观信息的减小而增加的情况只是对封闭系统而言。然而，问题在于，绝对封闭系统是不存在的。科学家们已计算出远达数十光年以外的恒星的移动引起的引力势能变化就可以使气体的微观状态发生实质性的变化。这样，从微观角度上看，孤立系统是不存在的，相空间的概率流体就不再是不可压缩的了。而且，这种扰动是随机的，故微观信息将被耗散。回到前述图 6-4-2 所示的实验，即使我们能满足 $t_0$ 时刻的全部初始条件，气体分子也不再可能如期望地那样全部回到 A 部，因为时过境迁，过去的历史已不能倒转了。所以，总结以上的分析可知，在实际的过程中，宏观信息不断地转化为微观信息，而微观信息又被随机的扰动所耗散，所有的过程，只有宏观信息的减少，宏观量熵的增加是确定不变的事实。

## 6.5 热寂学的批判

德国伟大的物理学家克劳修斯把从孤立系统得出的“熵增加原理”不恰当地推广到了具有开放系统特性的无限宇宙，得到了一个热寂世界，从而导致热寂学的诞生。这个结论，人们无论是从感性上还是理性上都难以接受，因此，自然地引发了各种争论。与此同时，又有不少学者对批判热寂学的基本论点也提出了疑问，认为热力学第二定律在未见事实反驳之前有理由期望它适用于无限的宇宙，理想状态下得出的结论也应能反映实际过程的规律，至于指责热寂学违反能量转换与守恒定律关于能量在质的方面也是守恒的——能量在质上消失了，实际上“批判”了热力学第二定律。

### 6.5.1 热寂学的观点

热寂学几乎从一开始就伴随熵概念和熵理论的诞生和发展。

**1. 热寂学的物理情景**

正是克劳修斯本人，将熵理论应用于宇宙，得出宇宙的熵值将趋于一个极大值 $S_{max}$，这样，宇宙将达到热平衡，进入热寂状态，整个宇宙趋于同一温度(接近零 K)。到那时，宇宙能量存在，但不可利用。这个结论的物理情景，引发了各种思考、议论和批判。

**2. 人们对热寂学的不同观点**

首先，人们认为不应该把宇宙看成是一个孤立的热力学系统。恩格斯(Friedrich Engels，1820—1895)称热寂学是归谬法。恩格斯向科学家们提出一个建议：“现在，现代自然科学必须从哲学那里采纳运动不灭的原理；它没有这个原理就不能继续存在。”并明确指

出,放射到太空中去的热,一定有可能通过某种途径转变为另一运动形式,在这种运动形式中,它将重新集结和活动起来。再者,有些科学家和哲学家为了回避宇宙起源、生存的理论和熵理论相联系热寂学,提出了用物质的不断消亡的解释来补充物质的创造,在宇宙中导入适当物质和能量,就可避免“热寂”,使宇宙形成稳恒态,宇宙将永恒地持续着。

我们知道,早在1946年,伽莫夫(Gamow,George,1904—1968)根据哈勃(Hubble,1889—1953)发现的红移现象和比利时天文学家和宇宙学家勒梅特(Lemaitre. Georges,1894—1966)的宇宙模型,已提出了一个“大爆炸”宇宙理论。按此理论推算,宇宙初期,密度减小,温度下降,并随着宇宙的膨胀,形成各种层次的结构,核子、原子核、原子、分子、地球、太阳系、银河系、星系图以及生命等。据此模型,我们今天的宇宙将遗留下一个2K微波辐射的背景。1964年两位美国射电天文学家发现了这个辐射背景,并证实为3K。由此,给予了“大爆炸”理论有力的实验验证。到70年代中期,大爆炸宇宙模型已成为天体物理学界公认的标准模型。

就是基于这个模型,有两种截然不同的认识:一种认为,若宇宙是静态的,则对每一静态系统,总有一个最大熵值 $S_{max}$,对于膨胀着的宇宙,每一瞬间熵可能达到的极大值 $S_{max}$ 是随着系统的膨胀与时俱增的,如果系统膨胀得足够快,系统不但不能每时每刻跟上进程以达到新的平衡,而且,实际上熵值 $S$ 的增长将落后于 $S_{max}$ 的增长,两者的差距则越来越大,虽然整个宇宙的熵不断增加,但它距离平衡态即热寂状态却越来越远,如图6-5-1所示。所以,将熵理论应用于宇宙将不会导致热寂学了。即使我们研究的宇宙不是静态的,那我们考查动态膨胀模型时,展现在我们面前的是一幅与热寂完全相反的景象:宇宙早期处于热平衡高温、高密度的状态,随着膨胀,逐渐出现温差,从无序向有序变化,这种变化并不能达到热寂,恰恰相反,它由此一步步发展出越来越复杂的多样化结构。在微观上形成核、原子、分子以致无机分子到高级的生物大分子;在宇观上形成星系团、星系、太阳、地球,以致在地球上演化出生物、人类及其由此组成的不断发展的社会。

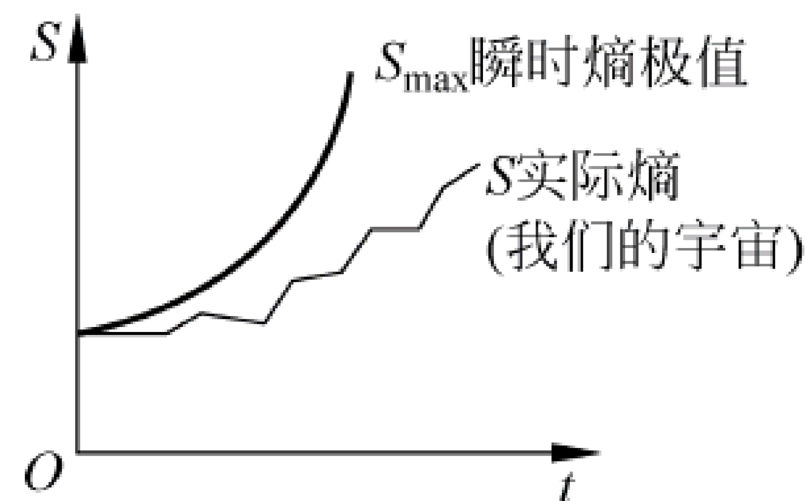


图6-5-1 实际熵 $S$ 的增长落后于实际熵值的最大值 $S_{max}$ 的增长

但另一种认识认为,20世纪60年代天文学家们在探测宇宙过去历史时期存在于银河的外层空间的放射源时发现,遥远的过去时期的放射源要比现在多得多。另外,类星体这类远离地球的遥远天体数量,也是现在较过去少得多。这些事实的发现,说明整个宇宙系统本身是在衰减。因此,宇宙,至少我们所能观察到的这部分宇宙将在遥远未来的某个时刻会回复到平衡状态而作为终结。

应当相信,不同理论观点上的争论,将进一步促进科学的进步。

### 6.5.2 热寂学的批判

热寂学是克劳修斯等人依据热力学第二定律提出的推论。它一出现就遭到恩格斯的严厉批判,斥之为“再也想不出比这更为愚蠢的东西了”。其基本论点可归为:第一,热寂学违背了能量守恒定律,“克劳修斯的第二原理不外乎是能消失了,如果不是在量上,那也是在质上消失了。”第二,宇宙达平衡态后须“外来的推动才会恢复”,如果承认宇宙的末日,就要承认它的开始,“就必须设想有上帝存在了”。

我国学者沿袭上述观点,物理化学和自然辩证法的教科书,公开发表的论文,只要论及

热寂学的，无不对之持批判态度。与恩格斯不同的是，我国学者是在确认热力学第二定律的前提下批判热寂学的。因而在援引和维护上述观点的同时，较多的是强调热寂学不是热力学第二定律的科学推论。如熵增加原理是在有限范围内得出的，不能无限推广到整个宇宙；熵增加原理适用于平衡态、孤立系统，而“平衡态，孤立系统等是一种理想状况，是一种科学抽象，只能在一个短的时间、小的范围内近似地符合，把它无限推广到一个无限的时间和空间中去显然是荒谬的”。我国学者具体是从三个方面对热寂学进行批判的。

**1. 不能把“熵增加原理”外推到“无限宇宙”**

(1) 熵增加原理虽然说明运动的“质”在有限时空内可以不守恒，但在无限宇宙内应该是守恒的，即宇宙内的排斥运动和吸引运动一定是互相平衡的；

(2) 功变热是排斥运动为主的运动，热变功的吸引运动是什么运动呢——恩格斯说，不是上帝，是未来自然科学的任务。

(3) 单纯的涨落理论缺乏说服力，即局部的吸引运动不足以抗衡排斥运动。

(4) 比利时物理学、化学家普里高金(Ilya Prigogine，1917—2003)的“耗散结构”理论指出，一个开放系统在远离平衡态的非线性区从混沌向有序转化的共同机制和规律——1977年获诺贝尔化学奖。这里所说的耗散结构是指需要与外界不断交换物质与能量的非孤立系统。

**2. 也不能把“熵增加原理”外推到“我们的宇宙”**

虽然，我们的宇宙的平均温度和平均密度在减小，但是我们的宇宙的温度分布、密度分布越来越不均匀，出现了无序向有序过渡的实际过程，不是孤立系统内熵增加原理所描述的过程。因此，我们的宇宙不能看成是孤立系统。

**3. “我们的宇宙”之未来也决不会“热死”——对立统一、盈则亏，否极泰来**

### 6.5.3　对热寂学批判的质疑

宇宙的热寂是一个无法验证的命题。它是不是热力学第二定律的合理推论，涉及自然科学的前沿、哲学中认识论、方法论的一些基本观点。纵观对热寂学的种种批判，学者又产生了一系列的疑问。

例如：“有限范围内得出的结论，不能推广到无限。”诚然，热力学第二定律是从有限范围得到的，但是“对自然界一切真实的认识都是对永恒的东西，对无限的东西的认识”，热力学第二定律作为自然界的基本规律，自然是真实的认识，是对无限的东西的认识，原则上它可以推广到整个宇宙。各种关于宇宙的理论，无一例外的都是从有限范围内得出而推广于无限的。当然，我们也看到不少学说、理论经验证，只在一定范围内适用。然而，这个适用与不适用的分界线，不是由已在多大范围内证实了来确定，而要由在什么范围出现事实反驳来确定。也就是说，某一基本规律如还没有发现它在什么范围不适用，我们就有理由期望它适用于无限。所以，如果我们不能证实热力学第二定律超过多大范围就不适用，仅因它是由有限而推广到无限，就认定热寂学是荒谬的，恐怕本身就违背了辩证唯物主义的认识论了。

由理想状况下得出的结论不能适用于实际过程吗？理论和实际固然有差异，但是，一切科学的抽象都更深刻、更真实地反映着自然。恩格斯就非常赞许卡诺研究卡诺循环的方法：卡诺“设计了一部理想的蒸汽机(或煤气机)，的确，这样一部机器就像几何学上的线和面一样是绝不可能制造出来的，但是它按照自己的方式起了像这些数学抽象所起的同样作用：它表现纯粹的、独立的和真正的过程”。正是这种理想的热机反映了实际热机的本质规律，正是对卡诺循环的研究，才引出熵的概念和热力学第二定律，它们的普遍适用性是自然科学

所公认的。可以说,运用理想状况或模型认识客观世界,不仅是有效的方法,而且是必经的途径。有时,理想状况和实际过程的差异会带来定量计算的误差,而定性地解决过程进行的方向趋势,通常总是行之有效的。那么,我们又有什么理由,仅因热力学第二定律中有理想化的概念,就认定由它推论得出的“热寂论”就是荒谬的呢?

至于承认宇宙热寂就要承论它的开始,由此而引出上帝的存在等说法,也是不能成立的。确实,宗教神学和各种唯心主义曾利用宇宙热寂说作为上帝创世说和世界末日论的依据。但这不是热寂学的过错。热寂学依据热力学规律和逻辑推理,宣判宇宙死刑,并没有征求上帝的意见,至于宇宙是如何诞生的,就更不容上帝插嘴了。宗教神学和唯心主义企图利用这一点证明上帝的存在,这不足为奇。任何新理论在对原有的未知做出解释的同时,都会引出新的未知,唯心主义都会用来证明上帝的存在。然而,斯宾诺莎说得好:“上帝=我不知道,但是无知不是论据。”

现在,宇宙大爆炸学说不仅提出了宇宙怎样诞生,甚至估算出宇宙诞生至今有150亿~200亿年。自然界万事万物都有一个发生、发展、灭亡的过程。为什么设想宇宙的灭亡就得承担窝藏上帝的罪责呢?“一切产生出的东西,都一定要灭亡。”(歌德《浮士德》)宇宙既在大爆炸中产生,就合理地要灭亡。当然,“热寂论”预言的灭亡不一定正确,它只是一家之言,可以争论,可以等待新的实验事实或新的理论来论证。但至少我们不应听说宇宙的灭亡就大惊失色,就认为是“离经叛道”而斥之为荒谬,归之于唯心。

热寂学的要害是能量转化中的守恒问题,即“热寂论”意味着在宇宙的演化中能量在量上守恒在质的变化上不可逆,即所谓“从质上消失了”,从而最终危及自然辩证法“运动的不灭性”这一根本命题。

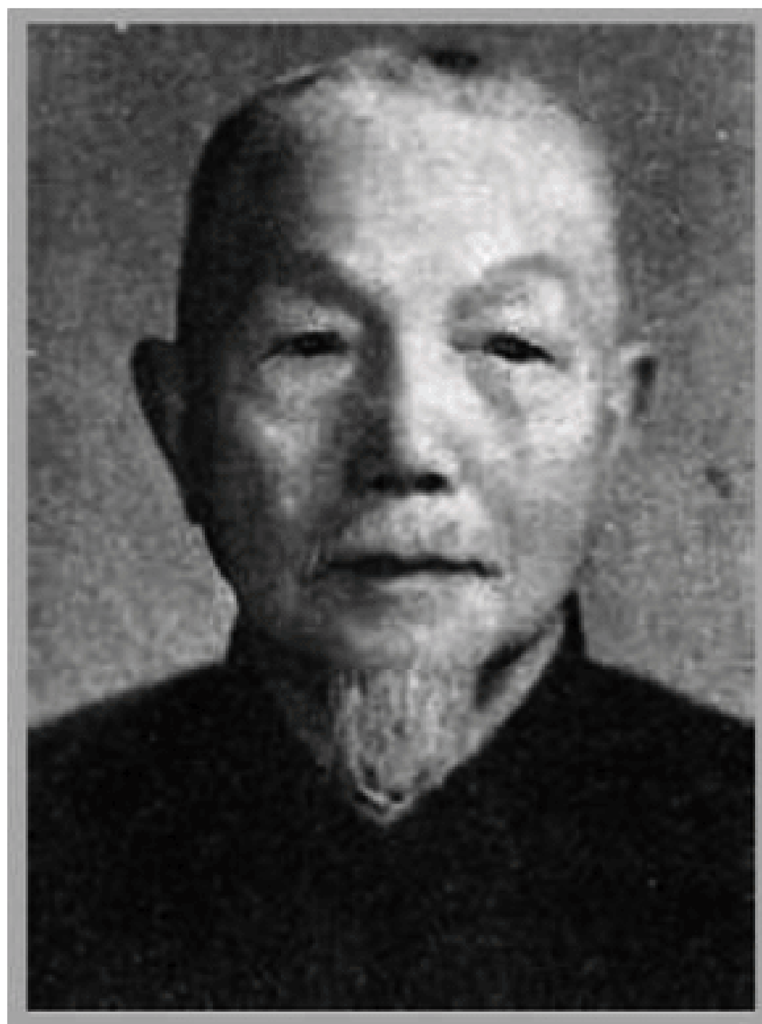
其实,即使达到热寂状态,运动也没有消失。热寂状态是平衡状态,在热力学中平衡从来就是动态平衡。在所谓热寂时,分子的热运动依然存在,则分子内部、原子内部的各种运动也依然存在。运动的不灭性是因能量守恒定律才得以最终确立的,而能量守恒定律即热力学第一定律其本来意义只是各种不同形式的能量相互转换时必在量上守恒,而所谓能量在质上守恒是一个未经证明的命题。违背热力学第一定律的过程不能发生,符合热力学第一定律的过程是否都能发生呢?这正是热力学第二定律所要解决的问题。众所周知,热力学第二定律可表述为“不能从单一热源取出热全部变为功而不留下其他变化”(第二类永动机未能成立)。而功(严格地说是机械能、化学能等)却可100%地转化为热能。所以,热和其他能量的转化在质上是不守恒的。热力学第二定律的其他表述也表明:自发过程总有做功能力的浪费或丧失(能量退化原理)。而对热寂学的批判恰恰将此当作热寂学的要害,故有理由认为:热寂学也许不是热力学第二定律的科学推论,但对热寂学的批判却实在是在“批判”热力学第二定律。

热力学第一定律、第二定律是自然界的基本定律。“自然界不是存在着,而是生成着并消逝着。”能量守恒只反映了自然界“存在着”,熵增加则是自然界“生成着并消逝着”在能量上留下的印记。我们批判热寂学时,强调热力学第二定律是从有限范围得出的结论,不能推广到无限的宇宙。而对同样范围产生的热力学第一定律,却认定它适用于整个宇宙,并将它扩展成不仅总量守恒,而且质上守恒。又因热寂学意味着能量“从质上消失了”而加以批判。但是,如上所述,能量在质上守恒即使在热力学适用的范围内也是不成立的。人们不禁疑问:对热寂学的批判成立吗?甚至连批判的论据都是不能自洽的。热寂学的真伪、热力学第二定律的适用范围还需等待未来广大学者和科学家去努力完成。

# 学习资料

## 一、中国近现代著名科学家(物理学)简介

葛正权(1896 年 1 月 18 日—1988 年 3 月 22 日),字秉衡,别号葛蠡,浙江省东阳县人。物理学家,南京大学物理系杰出校友。在精确验证麦克斯韦分子速度分布律方面取得国际公认的重大成就,是中国制氧工业先驱,创建中国第一个雷达研究所,开拓了我国雷达的研究和应用,在物理教学和应用方面,培养了许多人才。



他出生于浙江省东阳县礼村的一个农民家庭。祖辈世代勤劳务农,父亲半农半医,母亲纺纱织布以补家用。在勤俭和谐的家风熏陶下,葛正权从小养成刻苦、好学、爱家、爱乡的性格。他 5 岁进私塾读“四书”“五经”,光绪皇帝诏令“废除科举,兴办学校”,便转进小学读书。但几年的私塾求学,使他吸收了一些祖国的优秀文化,铸成了自尊、自爱和热爱祖国的优秀品格。他于 1920 年毕业于南京高等师范工科,留任助教,3 年后受聘于厦门大学。

1929 年他去美国洛杉矶南加州大学物理系,完成论文《短波干涉仪》获硕士学位,次年入伯克利加州大学,经过三年不懈努力用分子束测定 $Bi_2$ 分解热,并验证查克斯韦速度分布律成功,取得国际上领先的成果,获得博士学位。1933 年学成回国,受聘为武汉大学物理系教授,抗战时投笔从戎,任航空委员会第一氧气制造所所长,胜利后任空军制氧总厂厂长,国防部第六厅设计处长。1947 年建立雷达修理所,雷达研究所,打下了我国雷达事业的基础。1949 年南京政府溃逃台湾时,葛正权在战乱中把全部雷达设备和人才完整地保留下来,迎来了新中国的成立。1981 年主译近代物理学基础及其应用,上海科技出版社出版。1950 年他因病辞去雷达研究所所长职务,病愈后应聘为上海第二军医大学教授兼数理教研室主任,1984 年他的政治夙愿得偿,光荣地加入了中国共产党。直至 1988 年 3 月 22 日逝世于上海。享年 93 岁。综观他的一生,在一个混乱的历史时代,作为一名科学家,仍能以其所长,报效民族国家,未泯其爱国爱民之良知与良心,始终走在光明正确的道路上,是永远值得我们敬重的!

## 二、诺贝尔物理学奖简介

**1941 年未颁奖**。

**1942 年未颁奖**。

**1943 年诺贝尔物理学奖**。因发展分子束方法并测出质子磁矩。获奖者:斯特恩(Otto Stem,1888—1969)(美国)。评价:自旋是粒子的一种属性,存在一种角动量。但不是连续的只能取分立的几个值。用霍金的说法它表示粒子旋转几圈就会相同,例如自旋为 1 就是旋转 360°会变成一样(大多数物体都是如此,比如黑桃 A)。自旋为 2 的话就是旋转 180°变成一样(例如扑克牌中的方块)。自旋为 1/2 的话就得旋转两圈才能变成一样,可以参考莫比乌斯环带。斯特恩-革拉赫实验是物理史上的一个重要实验,它证明了自旋的存在。无论是物理理论和实际应用,这个实验都起到重要作用。后来经过探索他测出了质子和氘核的磁矩而且发现它们并不是人们理论推测的那样,说明质子可能还有内部结构。

**1944 年诺贝尔物理学奖**。发现测定原子核磁性的共振方法。获奖者：拉比(美国)。评价：自旋角动量只能是几个分立的值。而拉比在斯特恩-革拉赫实验的基础上测量出了让质子的角动量跃迁的方法(即往高的能级上跳)。而由此发展出的核磁共振技术相信我们大家都不会陌生。对人类健康起到了重要作用。

**1945 年诺贝尔物理学奖**。发现不相容原理。获奖者：W. 泡利(奥地利)。评价：1943 年和 1944 年诺贝尔物理学奖是奖给证明粒子有自旋存在的。其实在 1925 年泡利就提出了著名的不相容原理把自旋容纳到他的理论体系。其实这个理论是用来解释元素周期律的。我们曾被告知：原子的第一层轨道最多有 2 个电子，第二层最多有 8 个，第三层最多 18 个(即 $2n^2$)。这一切的原理全在泡利的不相容原理。泡利虽然是非常伟大的理论物理学家但实验却很糟糕，总是弄坏仪器。科学界有泡利效应的笑谈。据说有一次一个实验仪器不知道出了什么问题，后来又莫名其妙好了。人们百思不得其解，最后发现泡利那时正坐火车经过那座城市。后来人们就把这种现象称为泡利效应。

**1946 年诺贝尔物理学奖**。因发明产生很高压力的装置及利用这一装置在高压物理领域内所做出的贡献。获奖者：布里奇曼(美国)。评价：这次的物理学奖摆脱了量子力学那些深奥的东西，而变得朴实起来。也许你不知道布里奇曼是如何做到他的贡献，但你一看就知道他在干什么。说实话做物理实验就是要没事找事找一些一般不可能出现的极限状况，比如超高速之类的。而布里奇曼的高压正好帮助科学家们实现了一些极限情况，比如在高压下 100℃的冰。

**1947 年诺贝尔物理学奖**。因对大气高层物理性质的研究，特别是发现了阿普尔顿层。获奖者：E. V. 阿普尔顿(英国)。评价：今天我们的手机为什么可以打遍全球而畅通无阻，这要多亏了阿普尔顿的贡献。话说当年马可尼第一次实现了越洋通信时，人们就怀疑天空中存在电离层，不过难以证明。阿普尔顿实现了天空中的反射波和地面传播的电波的干涉现象(干涉现象要求比较苛刻，一般来说只有出自同一个波源)，证明了电离层的存在。后来又发现了反射能力更强的高空电离层。阿普尔顿的工作为环球无线电通信提供了重要的理论依据，从此无线电事业进入了一个新纪元。阿普尔顿还开辟了对电离层以及该层受太阳位置和日斑活动的影响的研究领域。

**1948 年诺贝尔物理学奖**。改进了威尔孙云雾室方法，并由此导致了在核物理领域和宇宙射线方面的一系列发现。获奖者：P. M. S. 布莱克特(英国)。评价：威尔孙发明云雾室，给物理学家以有力的武器观察大自然。不过初期的云雾室无法自动控制，拍照效率很低，基本上是瞎猫碰见死耗子，到布莱克特给云雾室增加了自动控制装置才使效率大大提高。

**1949 年诺贝尔物理学奖**。在核力理论工作的基础上预言了介子的存在。获奖者：汤川秀树(日本)。评价：根据量子力学的说法。电磁力是通过光子的作用产生的，那么原子核中的核力是否有类似的粒子呢？汤川秀树大胆推测认为这核力的交换是通过某种未知的粒子产生的，并且就算出这种粒子是电子质量的 200～300 倍，介于质子和电子之间(质子质量约为电子的 1700 倍)。这种粒子就称为介子。汤川秀树关于存在介子的预言是一个可与狄拉克的正电子预言相媲美的理论上的辉煌成就。从他的理论提出，到 p 介子的发现，标志着人类对物质的认识又向前跨进了一大步，即从认识原子核进入认识基本粒子的领域。值得一提的是汤川秀树是土生土长的日本制造，从未到欧美留学，酷爱庄子，可以说作为日本的第一位诺贝尔奖得主也是东亚文明的骄傲。

**1950 年诺贝尔物理学奖**。开发了用以研究核破坏过程的照相乳胶记录法并发现各种

介子。获奖者：C. F. 鲍威尔(英国)。评价：科学家在做实验时一般用乳胶来作记录，如果有粒子经过会留下痕迹。鲍威尔通过改善乳胶的灵敏度，取得了一系列重要成果，其中包括汤川秀树预言的 $\pi$ 介子以及 $\mu$ 介子。正是由于此鲍威尔开创了物理学的一个新的分支学科——粒子物理学，鲍威尔因此也被誉为粒子物理学之父。后来还有很多科学家按照鲍威尔的路线发现了新的粒子并获得诺贝尔物理学奖。

## 思考与练习

1. 简述温标、温度计和温度标度法的区别与联系。
2. 负的绝对温度的含义及其热力学性质。
3. 谈谈热力学第二定律的统计意义。
4. 根据热力学第二定律证明两条绝热线不能相交。
5. 根据熵增加原理证明第二定律的开尔文表述，从单一热源吸取热量使之完全变成有用的功而不引起其他变化是不可能的。
6. 从热力学第二定律的角度谈谈对“信息熵”概念的理解。
7. 从热力学第二定律的角度谈谈对“耗散结构”理论的看法。

# 第7章

# 电磁学中的场与路

麦克斯韦是一名伟大的理论物理学家，集电磁学之大成是他的杰出成就之一。麦克斯韦虽然不是实验物理学家，但是他运用自己卓越的理论思维能力和数学才能，在库仑、高斯、欧姆、安培、毕奥、萨伐尔、法拉第和奥斯特等人的一系列发现和实验成果的基础上，建立了完整的电磁场理论，即麦克斯韦方程组。麦克斯韦的杰出成就之一在于发现了交变电场可以产生（交变）磁场，从而预言了电磁波的存在，提出光的电磁学说，统一了电学、磁学和光学。

## 7.1　三种电场的比较

本节从非相对论的角度，对静电场、稳恒电场和涡旋电场这三种电场的物理意义、产生条件、场方程式及场的性质进行细致的比较讨论，从而对其场的特点、电场线、场强以及电势等有更深刻的理解。

### 7.1.1　三种电场的物理基础的比较

静止电荷产生的电场称为静电场。稳恒电流的电荷分布不随时间变化，这些作定向运动的电荷激发的电场亦不随时间变化，称为稳恒电场。变化的磁场产生的电场叫作感应电场，也叫作涡旋电场。在非相对论效应条件下，静电场、稳恒电场、涡旋电场这三种电场相互之间既有联系，又有区别。对三种电场的异同进行比较，有利于更好地把握这三种电场的性质及其物理内涵。加深对电磁学基本概念和理论体系的理解。

静电场、稳恒电场和涡旋电场具有以下三点共性：其一，它们都是电场，是物质的一种客观存在形式，都储存着电能；其二，它们对处于其中的电荷都有电场力的作用；其三，在这三种电场中移动电荷时，相应的电场力一般都要做功。

**1. 静电场的物理基础**

（1）库仑定律

库仑通过对著名的扭秤实验结果的分析，总结了两个静止点电荷间相互作用力的规律——库仑定律。其主要内容是：同号电荷相互排斥，异号电荷相互吸引；作用力沿两点电荷的连线；力的大小正比于每个点电荷电量的多少；力的大小反比于两点电荷之间距离的平方。

库仑定律具体表述如下：真空中静止点电荷 $q_1$ 对另一个静止点电荷 $q_2$ 的作用力 $F_{21}$ 为

$$\boldsymbol{F}_{21}=\frac{1}{4\pi\varepsilon_0}\frac{q_1q_2}{r^3}\boldsymbol{r}_{21} \tag{7-1-1}$$

式中 $\boldsymbol{r}_{21}$ 为由 $q_1$ 到 $q_2$ 的径矢，$\varepsilon_0$ 是真空介电常数。

库仑力满足牛顿第三定律，$\boldsymbol{F}_{12}=-\boldsymbol{F}_{21}=\dfrac{1}{4\pi\varepsilon_0}\dfrac{q_1q_2}{r^3}(-\boldsymbol{r}_{21})$，如图 7-1-1 所示。

（2）叠加原理

实验指出，两个点电荷间的作用力不因第三个电荷的存在而改变，不管一个体系中存在多少个点电荷，每一对点电荷之间的作用力都服从库仑定律，而任一点电荷所受到的力等于所有其他点电荷单独作用于该电荷的库仑力的矢量和，这一结论称为叠加原理。

根据力的叠加原理，各源电荷在 $P$ 点产生的电场的总场强为

$$\boldsymbol{E}=\boldsymbol{E}_1+\boldsymbol{E}_2+\cdots+\boldsymbol{E}_n=\sum_i\boldsymbol{E}_i \tag{7-1-2}$$

即一组点电荷共同产生的电场的场强等于每个点电荷在该点单独产生的电场的场强的矢量和。这一结论称为场强的叠加原理。

（3）总结

库仑定律和叠加原理是静电场最基本的规律。从它们出发可以得到环路定理和高斯定理。环路定理结合平方反比律，可得到电势和电荷分布的关系，高斯定理结合一定的对称性（有心力性质的反映），也能由电荷分布求出场强。它们之间的关系如图 7-1-2 所示。

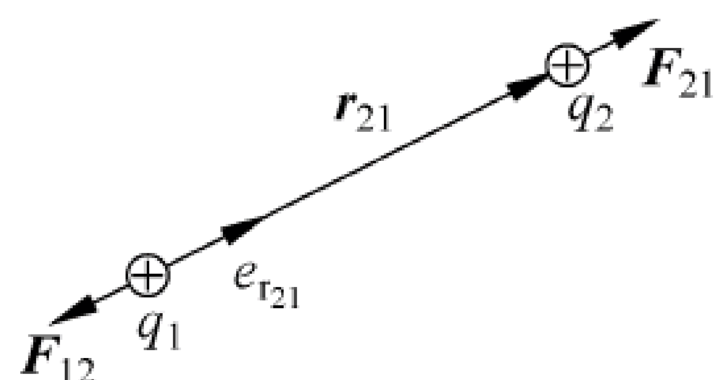



图 7-1-1　库仑力的图示

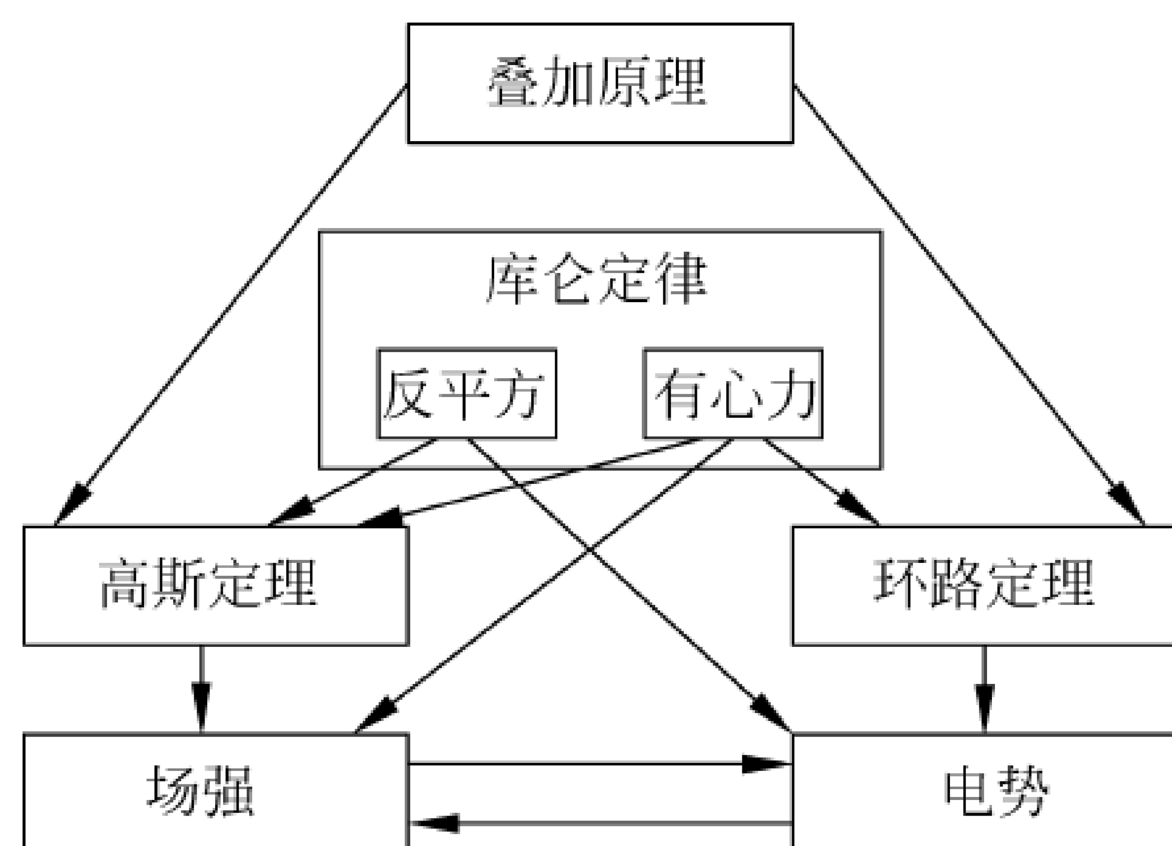



图 7-1-2　静电场规律之间的关系

**2. 稳恒电场的物理基础**

（1）电荷守恒定律及其数学表达——电流的连续性方程

在任何时刻，存在于孤立系统内部的正电荷与负电荷的代数和恒定不变。这一结论称为电荷守恒定律。电荷守恒定律是一切宏观过程和一切微观过程都必须遵循的基本规律。

设想在导体内任取一封闭曲面 $S$，则通过该封闭曲面的电流为

$$I=\oint_S\boldsymbol{J}\cdot\mathrm{d}\boldsymbol{s} \tag{7-1-3}$$

若 $I=\oint_S\boldsymbol{J}\cdot\mathrm{d}\boldsymbol{s}>0$，则表示有电荷通过封闭曲面向外迁移，单位时间内通过封闭曲面迁移的电量为 $I$。根据电荷守恒定律，单位时间内通过封闭曲面向外迁移的电量应等于该封闭曲面内单位时间所减少的电量。反之，若 $I=\oint_S\boldsymbol{J}\cdot\mathrm{d}\boldsymbol{s}<0$，则表示有电荷通过封闭曲面进入其内部，根据电荷守恒定律，单位时间内通过封闭曲面进入其内部的电量应等于该封闭曲面内单位时间所增加的电量。若以 $\mathrm{d}q/\mathrm{d}t$ 表示封闭曲面内的电量随时间的变化率，则有

$$\oint \boldsymbol{J} \cdot \mathrm{d}\boldsymbol{s} = -\frac{\mathrm{d}q}{\mathrm{d}t} \tag{7-1-4}$$

上式称为电流的连续性方程。它是电荷守恒定律的数学表述。

电流的连续性方程告诉我们,电流场的电流线是有头有尾的,凡有电流线发出的地方,那里的正电荷的量必随时间减少;凡有电流线汇聚的地方,那里的正电荷的量必随时间增加。上式也可写成

$$\oint_S \boldsymbol{J} \cdot \mathrm{d}\boldsymbol{s} = -\int_V \frac{\partial \rho}{\partial t} \mathrm{d}V \tag{7-1-5}$$

这是电荷守恒定律的积分形式。把上式中的面积分变为体积分,即可得电荷守恒定律的微分形式

$$\nabla \cdot \boldsymbol{J} + \frac{\partial \rho}{\partial t} = 0 \tag{7-1-6}$$

(2) 欧姆定律

当保持金属的温度恒定时,金属中的电流密度 $j$ 与该处的电场强度 $E$ 成正比,即

$$\boldsymbol{j} = \sigma \boldsymbol{E} \tag{7-1-7}$$

比例系数 $\sigma$ 称为金属的电导率。式(7-1-7)对大部分导体都是成立的,称为欧姆定律的微分形式。它反映了导体内部任一点的电流密度与该点的电场强度的联系。电流密度只决定于该点的电场强度与电导率,与其他地方的电场分布和电导率无关。

**3. 涡旋电场的实验基础——法拉第电磁感应定律**

法拉第发现当磁场发生变化时,附近闭合线圈中有电流通过,并由此总结出电磁感应定律:闭合线圈中的感应电动势与通过该线圈内部的磁通量变化率成正比。

设 $L$ 为闭合线圈,$S$ 为 $L$ 所围的一个曲面,$\mathrm{d}s$ 为 $S$ 上的一个面元。按照惯例,我们规定 $L$ 的围绕方向与 $\mathrm{d}s$ 的法线方向成右手螺旋关系。根据实验结果,当通过 $S$ 的磁通量增加时,在线圈 $L$ 上的感应电动势与我们规定的 $L$ 围绕方向相反,因此用负号表示。电磁感应定律表示为

$$\varepsilon = -\frac{\mathrm{d}}{\mathrm{d}t}\int_S \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{s} \tag{7-1-8}$$

**4. 小结**

静电场、稳恒电场和涡旋电场的实验基础和遵守基本定律如表 7-1-1 所示。

**表 7-1-1 三种电场实验基础的比较**

| 电场类型 | 实验基础 |
| --- | --- |
| 静电场 | 库仑定律　　　叠加原理 |
| 稳恒电场 | 电荷守恒定律　　欧姆定律 |
| 涡旋电场 | 法拉第电磁感应定律 |

## 7.1.2 三种电场产生条件的比较

**1. 稳恒电场产生的条件**

由于稳恒电流的电流密度不随时间变化,如果存在电流线发出或汇聚的地方,那么这些地方电荷的增加或减少的过程就将持续进行下去,这必将导致这些地方正电荷或负电荷的大量积聚,从而形成越来越强的电场,电场将阻碍电荷的继续积聚,电流将消失。如果要维

持电流恒定,就必须存在某种越来越强的非静电力以与由电荷积聚所形成的电场力相抵消,这在物理上是无法实现的,也是难以理解的。因而,对于真正的稳恒电流,必须不存在这种电荷不断积聚的地方,亦即对任何封闭曲面的通量必须等于零,即

$$\oint_S \boldsymbol{J} \cdot \mathrm{d}\boldsymbol{s} = 0 \tag{7-1-9}$$

这个结论也可以从电流的连续性方程直接得到。在恒定电流情况下,一切物理量不随时间而变,因而$\frac{\partial \rho}{\partial t}=0$,将其代入电流连续性方程$\nabla \cdot \boldsymbol{J}+\frac{\partial \rho}{\partial t}=0$,便可得到$\nabla \cdot \boldsymbol{J}=0$。

这意味着稳恒电流分布是无源的。任何时刻进入封闭曲面的电流线的条数与穿出该封闭曲面的电流线的条数相等,在电流场中既找不到电流线发出的地方,也找不到电流线汇聚的地方,稳恒电流的电流线只可能是闭合曲线,没有发源点,也没有终止点。这是稳恒电流的一个重要特性,称为稳恒电流的闭合性。通常也把它称为稳恒电流的连续性。

对于稳恒电流,电荷的定向运动具有下面的特点:在任何地点,其流失的电荷必被别处流来的电荷所补充,电荷的流动过程是空间每一点的一些电荷被另一些电荷代替的过程。正是这种代替,保证了电荷分布不随时间变化。分布不随时间变化的电荷产生的电场亦不随时间变化,这种电场称为稳恒电场,它是一种静态电场。

**2. 涡旋电场产生的条件**

考虑一个密绕的螺线环,当导线中通以电流时,电流的磁场都集中在环内,环外无磁场。今用一导线回路与环交链,根据法拉第电磁感应定律,当螺线环中的电流变化时,通过回路所圈围面积的磁感通量发生变化,回路中出现感应电流。感应电流的产生表明导体中自由电子受到某种非静电起源的作用力,因而产生定向运动。作用于电荷的力无非是电力和磁力两类。导线回路没有运动,而且它所在处连磁场都不存在,根本谈不上磁场的作用力问题。所以唯一的可能是导线中存在着电场。为了解释这一类电磁感应现象,麦克斯韦假设:除了电荷产生电场外,变化的磁场也产生电场。磁场变化在固定不动的导线回路中产生的感应电流,就是由变化的磁场产生的电场引起的。

变化的磁场产生的电场称为感应电场,也称为涡旋电场。由感应电场引起的电动势称为感生电动势。若用$\varepsilon$表示感应电场,则感生电动势为

$$\varepsilon = \oint_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} \tag{7-1-10}$$

根据法拉第电磁感应定律

$$\oint_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\frac{\mathrm{d}\Phi_m}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}\int_S \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{s} \tag{7-1-11}$$

式中$L$是任一闭合路径,它可以是某一导线回路,也可以是任一想象的闭合积分路径,$S$是以闭合路径$L$为周界的任意曲面。

由于回路是固定不动的,故上式可写成

$$\oint_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\int_S \frac{\partial \boldsymbol{B}}{\partial t} \cdot \mathrm{d}\boldsymbol{s} \tag{7-1-12}$$

即感应电场对任意闭合路径的线积分取决于磁感强度的变化率对这一闭合路径所圈围面积的通量。由式(7-1-12)可以看出,涡旋电场存在的前提是$\partial \boldsymbol{B}/\partial t \neq 0$。

**3. 小结**

表7-1-2给出三种电场产生条件的比较。

表 7-1-2 三种电场产生条件的比较

| 电场类型 | 产生条件 |
|---|---|
| 静电场 | 静止电荷 |
| 稳恒电场 | $\nabla \cdot \boldsymbol{J}=0$ |
| 涡旋电场 | $\partial \boldsymbol{B}/\partial t \neq 0$ |

## 7.1.3 三种电场的场方程式及场性质的比较

### 1. 静电场的场方程式及场的性质

(1) 静电场是有源场

电场对任意封闭曲面的电通量只决定于被包围在封闭曲面内部的电荷，且等于包围在封闭曲面内电量代数和除以曲面面积，与封闭曲面外的电荷无关。这一结论就是静电场的高斯定理。

$$\oint_S \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{s} = \frac{1}{\varepsilon_0}\sum_i q_i \tag{7-1-13}$$

上式为高斯定理的积分形式。式中 $E$ 是空间的总场强，$S$ 为任意形状的封闭曲面，则应理解为包围在 $S$ 内的总电量，即电量的代数和。

若包围在 $S$ 面内的电荷具有一定的体分布，电荷体密度为 $\rho$，则高斯定理又可改写为

$$\oint_S \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{s} = \frac{1}{\varepsilon_0}\int_V \rho \mathrm{d}v \tag{7-1-14}$$

式中 $V$ 是 $S$ 所包围的体积。将上式中的体积不断缩小，根据矢量场散度的定义，式子左边趋于电场 $E$ 的散度乘上体积元的 $\mathrm{d}v$，而右边趋于$\frac{1}{\varepsilon_0}\rho \mathrm{d}v$，因此有

$$\nabla \cdot \boldsymbol{E} = \frac{\rho}{\varepsilon_0} \tag{7-1-15}$$

这就是高斯定理的微分形式。它表明电荷是电场的源，电场线从正电荷发出而终止于负电荷。高斯定理反映了静电场是有源场这一特性。

(2) 静电场是无旋场

静电场的场强 $E$ 沿任一闭合路径 $L$ 的线积分为零，即静电场的环流为零。这一结论称为静电场的环路定理。

$$\oint_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = 0 \tag{7-1-16}$$

把回路 $L$ 不断缩小，使它包围着一个面元 $\mathrm{d}s$。根据旋度的定义，上式左边趋于$\nabla \times \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{s}$，由 $\mathrm{d}s$ 的任意性得

$$\nabla \times \boldsymbol{E} = 0 \tag{7-1-17}$$

式(7-1-17)表明静电场是无旋场。

(3) 静电场是保守场

环流为零的场为保守场。由上可知静电场的环流为零，所以静电场是保守场。

设 $L_1$ 和 $L_2$ 为由 $P_1$ 点到 $P_2$ 点的两条不同路径，$L_1$ 与 $-L_2$ 合成闭合回路，因此 $\int_{L_1} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} - \int_{L_2} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = 0$，即

$$\int_{L_1} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = \int_{L_2} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} \tag{7-1-18}$$

上式表明电荷由点 $P_1$ 移至点 $P_2$ 时电场对它所做的功与路径无关，而只和两端点有关，即在静电场中经过不同的路径将点电荷从一处移到另一处时，电场力对电荷所做的功是完全一样的。

**2. 稳恒电场的场方程式及场的性质**

稳恒电场的场方程式可以表示为

$$\boldsymbol{J}=\sigma\boldsymbol{E} \tag{7-1-19}$$

在稳恒电流情况下尽管电荷在流动，但由于电荷分布不随时间变化，所以从这一点看，由稳定流动的电荷所激发的稳恒电场在本质上与由静止电荷所激发的静电场是相同的。稳恒电场同样遵守高斯定理和环路定理，因此，稳恒电场也是保守力场。

**3. 涡旋电场的场方程式及场的性质**

由 $\oint_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=-\int_S \frac{\partial \boldsymbol{B}}{\partial t}\cdot \mathrm{d}\boldsymbol{s}$ 可知，涡旋电场的旋度并不为零。这表明涡旋电场是有旋场。这是涡旋电场与静电场的一个很大的区别。涡旋电场的性质与稳恒电流的磁场的性质非常相似。如果说，电流是磁场的涡旋中心的，那么变化的磁场就是涡旋电场的涡旋中心。由于自然界中不存在磁荷，故磁场是无源场，磁感线是闭合的。在只有感应电场分布的空间亦无电荷存在，因而涡旋电场也是无源场。涡旋电场的环流不为零，因此涡旋电场是非保守场。

与稳恒电流的磁场的基本方程式相似，感应电场的基本方程为

$$\begin{cases}\oint_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=-\int_S \frac{\partial \boldsymbol{B}}{\partial t}\cdot \mathrm{d}\boldsymbol{s}\\ \oint_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{s}=0\end{cases} \tag{7-1-20}$$

写成微分形式得

$$\begin{cases}\nabla\times \boldsymbol{E}=-\frac{\partial \boldsymbol{B}}{\partial t}\\ \nabla\cdot \boldsymbol{E}=0\end{cases} \tag{7-1-21}$$

**4. 小结**

表 7-1-3 列出了三种电场的场方程式及场的性质的比较。

**表 7-1-3 三种电场的场方程式及场的性质的比较**

| 电场类型 | 场 性 质 | | | 场 方 程 式 | |
|---|---|---|---|---|---|
| | 有无源 | 有无旋 | 是否保守场 | 积分形式 | 微分形式 |
| 静电场 | 有源场 | 无旋场 | 保守场 | $\oint_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{s}=\frac{1}{\varepsilon_0}\int_V \rho \mathrm{d}v$<br>$\oint_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=0$ | $\nabla\cdot \boldsymbol{E}=\frac{\rho}{\varepsilon_0}$<br>$\nabla\times \boldsymbol{E}=0$ |
| 稳恒电场 | 有源场 | 无旋场 | 保守场 | $\boldsymbol{J}=\sigma\boldsymbol{E}$ | |
| 涡旋电场 | 无源场 | 有旋场 | 非保守场 | $\oint_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=-\int_S \frac{\partial \boldsymbol{B}}{\partial t}\cdot \mathrm{d}\boldsymbol{s}$<br>$\oint_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{s}=0$ | $\nabla\times \boldsymbol{E}=-\frac{\partial \boldsymbol{B}}{\partial t}$<br>$\nabla\cdot \boldsymbol{E}=0$ |

### 7.1.4 三种电场的物理图像——电场线的比较

静电场的物理图像：电荷是电场的源，电场线从正电荷发出而终止于负电荷，在自由空间中电场线连续通过而不会中断；在静电情形下电场没有旋涡状结构，如图 7-1-3 所示。

稳恒电场的电场线的性质与静电场的电场线的性质亦完全相同。电场中用等势线描绘这一实验就是利用这一等价性，用稳恒电场来模拟静电场。

涡旋电场的物理图像：在变化的磁场周围有一圈圈闭合的涡旋电场线围绕着，呈旋涡状，涡旋电场也因此而得名。变化的磁场就是涡旋电场的涡旋中心。电场强度 $\boldsymbol{E}$ 与磁感强度的变化率$\dfrac{\partial \boldsymbol{B}}{\partial t}$组成左手螺旋，如图 7-1-4 所示。

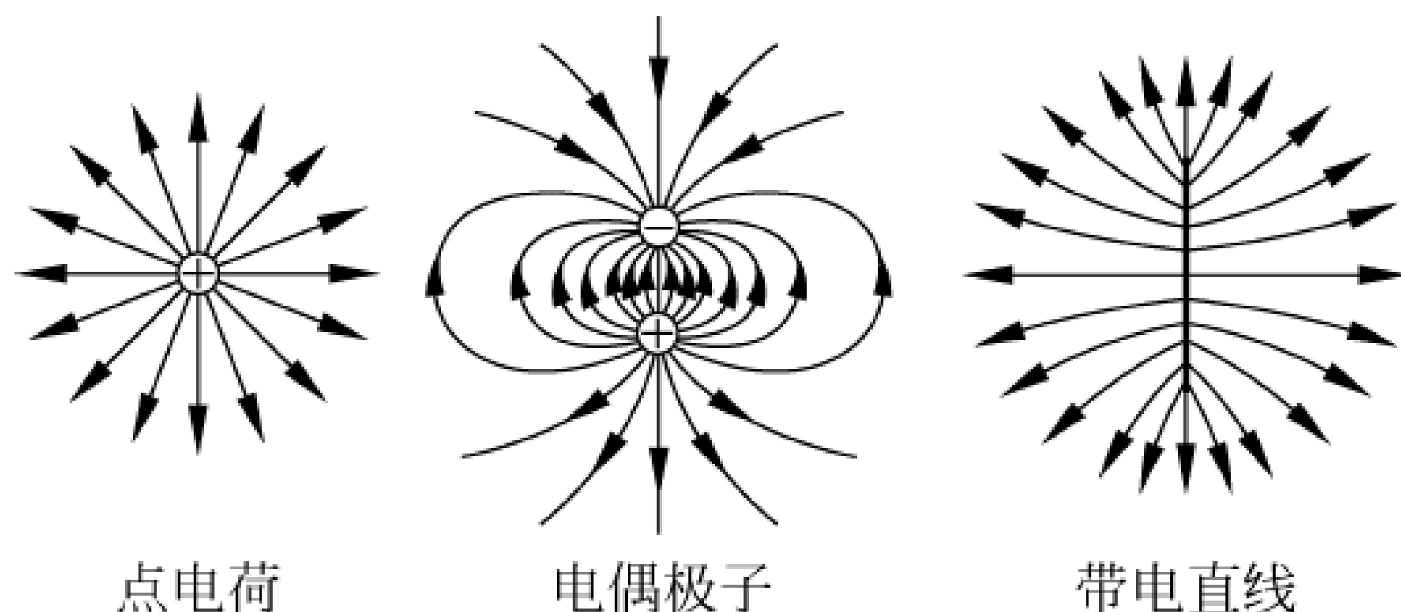



图 7-1-3 静电场的电场线

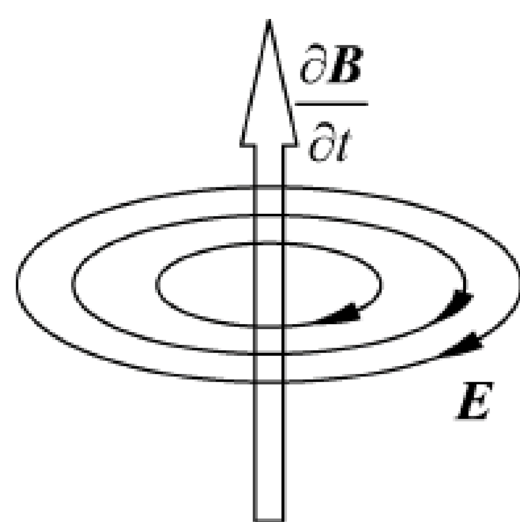



图 7-1-4 涡旋电场的电场线

### 7.1.5 三种电场的场强公式的比较

若电量为 $q_0$ 的试探电荷在场内某点受到的作用力为 $\boldsymbol{F}$，则该点的电场强度定义为

$$\boldsymbol{E}=\frac{\boldsymbol{F}}{q_0} \tag{7-1-22}$$

即电场内任意一点的电场强度等于一个单位电量的电荷在该点受到的作用力，电场强度的方向与正电荷在该点受力的方向相同。

**1. 静电场的场强公式**

设电荷连续分布于区域 $V$ 内。在 $V$ 内某点 $x'$ 上取一个体积元 $\mathrm{d}v'$，在 $\mathrm{d}v'$ 内所含的电荷 $\mathrm{d}Q$ 等于该点上的电荷密度 $\rho(x')$ 乘以体积 $\mathrm{d}v'$，即 $\mathrm{d}Q=\rho(x')\mathrm{d}v'$。设由源点 $x'$ 到场点 $x$ 的距离为 $\boldsymbol{r}$，则 $P$ 点上的电场强度 $E$ 为

$$\boldsymbol{E}(x)=\int\frac{\rho(x')\boldsymbol{r}}{4\pi\varepsilon_0 r^3}\mathrm{d}v' \tag{7-1-23}$$

式中积分遍及电荷分布区域。式(7-1-23)即为静电场的电场强度分布的积分形式。

**2. 稳恒电场的场强公式**

由欧姆定律 $\boldsymbol{j}=\sigma\boldsymbol{E}$ 变形可以得到稳恒电场的场强公式：

$$\boldsymbol{E}=\boldsymbol{j}/\sigma \tag{7-1-24}$$

**3. 涡旋电场的场强公式**

将 $\oint_L \boldsymbol{E}\cdot\mathrm{d}\boldsymbol{l}=-\int_S\dfrac{\partial\boldsymbol{B}}{\partial t}\cdot\mathrm{d}\boldsymbol{s}$ 化为微分形式后得到

$$\nabla\times\boldsymbol{E}=-\frac{\partial\boldsymbol{B}}{\partial t} \tag{7-1-25}$$

**4. 小结**

表 7-1-4 给出三种电场的场强公式的比较。

**表 7-1-4　三种电场的场强公式的比较**

| 共同的定义式 | 静　电　场 | 稳 恒 电 场 | 涡 旋 电 场 |
|---|---|---|---|
| $\boldsymbol{E}=\dfrac{\boldsymbol{F}}{q_0}$ | $\boldsymbol{E}(x)=\int\dfrac{\rho(x')\boldsymbol{r}}{4\pi\varepsilon_0 r^3}\mathrm{d}v'$ | $\boldsymbol{E}=\dfrac{\boldsymbol{j}}{\sigma}$ | $\nabla\times\boldsymbol{E}=-\dfrac{\partial\boldsymbol{B}}{\partial t}$ |

## 7.1.6　三种电场的电势的比较

静电场是保守场，因此可以引入电势的概念。把单位正电荷由点 $P_1$ 移至点 $P_2$，电场 $E$ 对它所做的功为 $\int_{P_1}^{P_2}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{l}$，定义为点 $P_1$ 和点 $P_2$ 的电势差。若电场对电荷做了正功，则电势 $\varphi$ 下降。所以有

$$\varphi(P_2)-\varphi(P_1)=-\int_{P_1}^{P_2}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{l} \tag{7-1-26}$$

通常选无穷远处为参考点，令 $\varphi(\infty)=0$，则有

$$\varphi(P)=\int_{P}^{\infty}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{l} \tag{7-1-27}$$

式为静电场的电势公式。

稳恒电场也是保守场，因此也可以引入电势进行描述；涡旋电场是非保守场，因而没有电势这个概念。表 7-1-5 列出三种电场的电势的比较。

**表 7-1-5　三种电场的电势的比较**

| 电场类型 | 电　　势 |
|---|---|
| 静电场 | $\varphi(P)=\int_{P}^{\infty}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{l}$ |
| 稳恒电场 | $\varphi(P)=\int_{P}^{\infty}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{l}$ |
| 涡旋电场 | 不存在电势这个概念 |

## 7.1.7　结论

根据已对三种电场进行的细致比较，可以将其主要内容归纳为表 7-1-6。

**表 7-1-6　三种电场的比较**

| | 静　电　场 | 稳 恒 电 场 | 涡 旋 电 场 |
|---|---|---|---|
| 共同点 | (1) 它们都是电场，是物质的一种客观存在形式，都储存着电能 | | |
| | (2) 它们对处于其中的电荷都有电场力的作用 | | |
| | (3) 在这三种电场中移动电荷时相应的电场力一般都要做功 | | |
| 物理基础 | 库仑定律<br>叠加原理 | 电荷守恒定律<br>欧姆定律 | 法拉第电磁感应定律 |
| 产生条件 | 静止电荷 | $\nabla\cdot\boldsymbol{J}=0$ | $\partial\boldsymbol{B}/\partial t\neq 0$ |

续表

| | | 静电场 | 稳恒电场 | 涡旋电场 |
|---|---|---|---|---|
| 场性质 | 有无源 | 有源场 | 同左列 | 无源场 |
| | 有无旋 | 无旋场 | | 有旋场 |
| | 是否保守场 | 保守场 | | 非保守场 |
| 场方程式 | 积分形式 | $\oint_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{s}=\frac{1}{\varepsilon_0}\int_V \rho \mathrm{d}v$<br>$\oint_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=0$ | | $\oint_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=-\int_S \frac{\partial \boldsymbol{B}}{\partial t}\cdot \mathrm{d}\boldsymbol{s}$<br>$\oint_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{s}=0$ |
| | 微分形式 | $\nabla\cdot\boldsymbol{E}=\frac{\rho}{\varepsilon_0}$<br>$\nabla\times\boldsymbol{E}=0$ | | $\nabla\cdot\boldsymbol{E}=0$<br>$\nabla\times\boldsymbol{E}=-\frac{\partial \boldsymbol{B}}{\partial t}$ |
| 电场线 | | 电场线从正电荷发出而终止于负电荷,在自由空间中电场线连续通过而不会中断 | | 旋涡状的闭合曲线;变化的磁场是涡旋电场的涡旋中心 |
| 场强公式 | $\boldsymbol{E}=\frac{\boldsymbol{F}}{q_0}$ | $\boldsymbol{E}(x)=\int \frac{\rho(x')\boldsymbol{r}}{4\pi\varepsilon_0 r^3}\mathrm{d}v'$ | $\boldsymbol{E}=\frac{\boldsymbol{j}}{\sigma}$ | $\nabla\times\boldsymbol{E}=-\frac{\partial \boldsymbol{B}}{\partial t}$ |
| 电势 | | $\varphi(P)=\int_P^\infty \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}$ | 同左列 | 不存在电势这个概念 |

## 7.2 位移电流及其热效应

本节通过对电介质及其极化特性的讨论,分析了在电介质极化中产生位移电流的微观机制,肯定了位移电流中束缚电荷电流热效应的存在,并探讨了其热效应的主要特征。

位移电流热效应存在与否及其存在的本质和形式应与位移电流所处空间介质的性质有关。真空中的位移电流源于变化的电场,介质中的位移电流既有变化电场的部分,又有介质极化电荷的运动。介质的极化,既有束缚电荷运动而形成的等效于传导电流性质的束缚电荷电流,又有伴随着介质内部由于电偶极子(极化场不太强)的形成和转向而发生的结构变化,这种变化会引起电介质的一些性质(如热力学性质等)发生改变,改变的程度与电介质的类型有关,当外极化场较强及其频率较高时,介质内还会产生一些非线性效应。因此,其位移电流的热效应问题就会因不同情形而由本质不同的几种部分组成。本文就外极化场不太强即介质电极化强度矢量与极化电场强度成线性关系的情形,根据电介质的类型,对其位移电流的热效应问题作一初步探讨。

### 7.2.1 位移电流的产生机制和规律

由于静电场中的高斯定理在迅变场中仍然成立,将其应用于图 7-2-1 中由 $S_1$ 和 $S_2$ 构成的闭合曲面 $S$ 上,得

$$\oiint_S \boldsymbol{D}\cdot \mathrm{d}\boldsymbol{s}=q \tag{7-2-1}$$

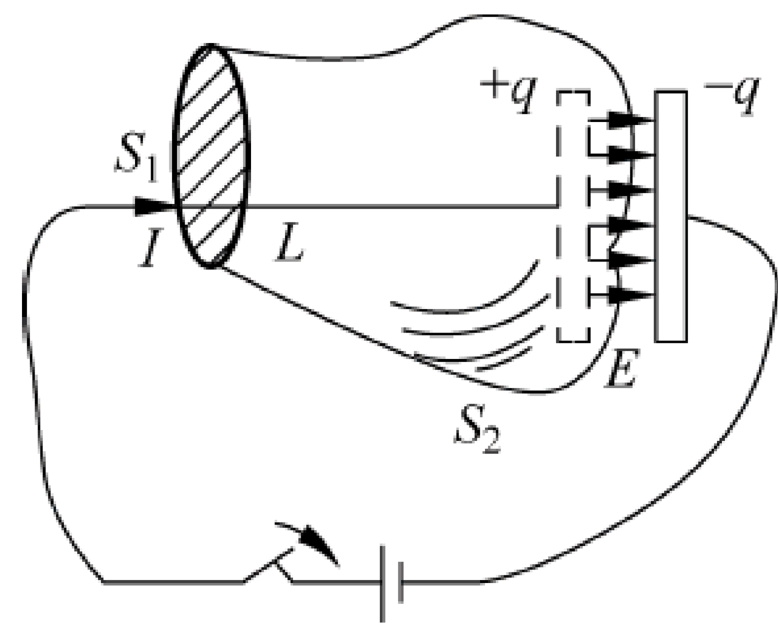


图 7-2-1 涡旋电场的电场线

将式(7-2-1)代入电流连续性方程 $\oiint_S \boldsymbol{J}\cdot \mathrm{d}\boldsymbol{s}=-\frac{\mathrm{d}q}{\mathrm{d}t}$ 并假定 $S$

不随时间变化，有

$$\oiint_S \boldsymbol{J}_{传} \cdot \mathrm{d}\boldsymbol{s} = -\frac{\mathrm{d}q}{\mathrm{d}t} = -\frac{\partial}{\partial t}\oiint_S \boldsymbol{D} \cdot \mathrm{d}\boldsymbol{s} = -\oiint_S \frac{\partial \boldsymbol{D}}{\partial t} \cdot \mathrm{d}\boldsymbol{s} \tag{7-2-2}$$

移项得

$$\oiint_S \left(\boldsymbol{J}_{传} + \frac{\partial \boldsymbol{D}}{\partial t}\right) \cdot \mathrm{d}\boldsymbol{s} = 0 \tag{7-2-3}$$

或 $\iint_{S1}\left(\boldsymbol{J}_{传} + \frac{\partial \boldsymbol{D}}{\partial t}\right) \cdot \mathrm{d}\boldsymbol{s}_1 + \iint_{S2}\left(J_{传} + \frac{\partial \boldsymbol{D}}{\partial t}\right) \cdot \mathrm{d}\boldsymbol{s}_2 = 0$，因 $S = S_1 + S_2$ 是任一边界 $L$ 的封闭曲面，则应有 $\mathrm{d}\boldsymbol{s}_1 = \mathrm{d}\boldsymbol{s}_2$，故可得

$$\iint_{S1}\left(\boldsymbol{J}_{传} + \frac{\partial \boldsymbol{D}}{\partial t}\right) \cdot \mathrm{d}\boldsymbol{s} = \iint_{S_2}\left(\boldsymbol{J}_{传} + \frac{\partial \boldsymbol{D}}{\partial t}\right) \cdot \mathrm{d}\boldsymbol{s} \tag{7-2-4}$$

进一步分析知，在图 7-2-1 中，$S_1$ 上有$\frac{\partial \boldsymbol{D}}{\partial t} = 0$，$S_2$ 上有 $\boldsymbol{J}_{传} = 0$，则式(7-2-4)变为

$$\iint_{S1} \boldsymbol{J}_{传} \cdot \mathrm{d}\boldsymbol{s} = \iint_{S2} \frac{\partial \boldsymbol{D}}{\partial t} \cdot \mathrm{d}\boldsymbol{s} \tag{7-2-5}$$

式(7-2-5)表明：电位移矢量 $\boldsymbol{D}$ 随时间的变化率等效于真实的电荷流动，即等效于电流。因此麦克斯韦把 $\iint_{S2} \frac{\partial \boldsymbol{D}}{\partial t} \cdot \mathrm{d}\boldsymbol{s}$ 称为位移电流，$\frac{\partial \boldsymbol{D}}{\partial t}$ 是位移电流密度，并将传导电流 $I = \iint_{S_1} \boldsymbol{J}_{传} \cdot \mathrm{d}\boldsymbol{s}$ 与位移电流一起称为全电流。在介质空间，如果传导电流与位移电流同时存在，则其周围空间磁场由二者共同产生，安培环路定理应修改为

$$\oint_L \boldsymbol{H} \cdot \mathrm{d}\boldsymbol{l} = \iint_{S1}\left(\boldsymbol{J}_{传} + \frac{\partial \boldsymbol{D}}{\partial t}\right) \cdot \mathrm{d}\boldsymbol{s} = \iint_{S2}\left(\boldsymbol{J}_{传} + \frac{\partial \boldsymbol{D}}{\partial t}\right) \cdot \mathrm{d}\boldsymbol{s} = I_{传} + I_{位} \tag{7-2-6}$$

由此便消除了非稳态下电流不连续的矛盾。

真空中位移电流的意义在于揭示了电与磁相互转化本质的另一个重要方面：即真空中没有真实电荷移动而仅有变化的电场产生的位移电流，与传导电流一样，按同一规律在其周围空间激发出涡旋磁场，即变化的电场也是涡旋磁场的源。

### 7.2.2　束缚电荷电流的产生机制和规律

由位移电流 $I_{\mathrm{D}} = \iint_S \boldsymbol{J}_{\mathrm{D}} \cdot \mathrm{d}\boldsymbol{s} = \iint_S \frac{\partial \boldsymbol{D}}{\partial t} \cdot \mathrm{d}\boldsymbol{s} = \iint_S \left(\varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} + \frac{\partial \boldsymbol{P}}{\partial t}\right) \cdot \mathrm{d}\boldsymbol{s}$ 知，其与传导电流的区别在于：在真空中，电极化强度矢量 $\boldsymbol{P} = 0$，且$\frac{\partial \boldsymbol{D}}{\partial t} = 0$，则 $\boldsymbol{J}_{\mathrm{D}} = \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}$，无电荷的定向移动；在电介质中，位移电流密度应分为两部分，$\varepsilon_0 \frac{\partial \boldsymbol{D}}{\partial t}$ 与真空中的情况完全相同，另一部分即电极化强度矢量 $\boldsymbol{P}$ 随时间的变化率$\frac{\partial \boldsymbol{P}}{\partial t}$，是由于极化时，束缚电荷的移动而形成的，一般文献称之为束缚电荷的电流或束缚电流。

束缚电荷电流源于电介质的极化。电介质极化有三种基本的极化过程，即无极分子的感生(或电子)极化；有极分子的转向极化；离子晶体的离子极化。转向极化也伴有感生极化。

在一般情况下，对于一定大小的外电场强度，极化强度矢量 $\boldsymbol{P}$ 的数值和位相是外场频

率的函数，因为在不同频率下会发生不同的极化过程，而且还伴有共振极化发生。在各向同性介质中，极化强度矢量与外电场强度的关系为

$$\boldsymbol{P}=\varepsilon_0 x_e \boldsymbol{E}=\varepsilon_0(\varepsilon_r-1)\boldsymbol{E} \tag{7-2-7}$$

实验表明：极化率 $x_e$ 也可分为三部分：由于电子壳层相对于核的位移所产生的感生（电子）极化率；具有永久电偶极矩的诸分子所贡献的偶极子极化率；离子晶体中正负离子相对位移的离子极化率。

光频下的介电常数几乎完全由感生极化率确定，因为在高频下，由于分子和离子的惯性，偶极子极化率和离子极化率贡献很小；在光频段，如果外光电场很强，比如强激光电场强度作用，则介质的极化强度矢量 $\boldsymbol{P}$ 与入射光射场强之间不再是简单的线性关系，即式(7-2-7)不成立，而是与光电场的二次项或高次项有关，这时介质会产生诸如混频、自聚焦和自感应衰减等非线性光学现象。

各向异性电介质在不同方向的极化率是不同的，比如晶状固体在不同晶轴方向上有不同的介电性质，其原因是组成晶体原子的电荷在某些方向的运动比其他方向“容易”，这时，当外场沿某几个从优方向极化时，电极化强度矢量 $\boldsymbol{P}$ 才会与 $\boldsymbol{E}$ 平行。

但是，不管电介质的结构和性质如何，也不论外场如何变化，电介质中的极化强度矢量 $\boldsymbol{P}=\dfrac{\sum \boldsymbol{p}}{\Delta V}$ 随时间而改变 $\dfrac{\partial \boldsymbol{P}}{\partial t}=\dfrac{\sum \boldsymbol{p}}{\partial t}$ 的地方都会有电流，而且是束缚电荷的真正运动。设电介质内单位体积中有 $N$ 个偶极子，在时间间隔 $\mathrm{d}t$ 内，每一个偶极子的电偶极矩从 $\boldsymbol{P}$ 改变到 $\boldsymbol{P}+\mathrm{d}\boldsymbol{P}$，则宏观极化强度矢量 $\boldsymbol{P}=N\boldsymbol{p}$ 改变到 $\boldsymbol{P}+\mathrm{d}\boldsymbol{P}=N(\boldsymbol{p}+\mathrm{d}\boldsymbol{p})$。假设 $\mathrm{d}\boldsymbol{p}$ 是由于一个电荷 $q$ 移动 $\mathrm{d}l$ 而引起的，在介质中每个分子、原子或离子都有 $q\mathrm{d}l=\mathrm{d}p$，则在时间 $\mathrm{d}t$ 内，确实有一个密度为 $\rho_P=Nq$ 的电荷云以速度 $\boldsymbol{v}=\dfrac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}t}$ 移动，它等效于一个电流密度为 $\boldsymbol{J}_P$ 的传导电流，即

$$\boldsymbol{J}_P=\rho_P\boldsymbol{v}=Nq\frac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}t}=N\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}=\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} \tag{7-2-8}$$

此关系式与上述极化机制、过程和具体模型的细节无关，极化强度矢量 $\boldsymbol{P}$ 随时间的变化率 $\dfrac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}$ 等效于传导电流，本质上是真实的电荷移动，从这个意义上讲，它与传导电流没有本质的区别。只是通常的传导电流是自由电荷的移动，这里的 $\boldsymbol{J}_P=\dfrac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}=Nq\dfrac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}t}$ 是束缚电荷的移动，而且一般不会有稳恒的束缚电流。

### 7.2.3 位移电流的热效应问题

引入束缚电荷的电流后，一方面对一般电介质中位移电流的本质和构成会有更加清晰的认识，由麦克斯韦方程有

$$\nabla\times\boldsymbol{H}=\boldsymbol{J}+\frac{\partial \boldsymbol{D}}{\partial t}=\boldsymbol{J}+\frac{\partial(\varepsilon_0\boldsymbol{E}+\boldsymbol{P})}{\partial t} \tag{7-2-9}$$

式(7-2-9)中 $\boldsymbol{J}$ 是传导电流密度，$\dfrac{\partial \boldsymbol{D}}{\partial t}=\varepsilon_0\dfrac{\partial \boldsymbol{E}}{\partial t}+\dfrac{\partial \boldsymbol{P}}{\partial t}$ 被麦克斯韦称为位移电流密度。在真空中 $\boldsymbol{P}=0$，$\boldsymbol{D}=\varepsilon_0\boldsymbol{E}$，则 $\boldsymbol{J}_{位}=\dfrac{\partial \boldsymbol{D}}{\partial t}=\varepsilon_0\dfrac{\partial \boldsymbol{E}}{\partial t}$ 才是无真实电荷移动但激发磁场与传导电流等效的真

正意义上的位移电流。由于没有真实的电荷移动,因此也就没有传统意义上的热效应,这是真空位移电流与传导电流最为本质的区别。因为焦耳热的本质是导体中自由电子或移动的带电粒子受电场力作用而发生定向移动时,不断与金属晶格发生碰撞而把其定向动能的一部分传给晶格而使晶格围绕其平衡位置的振动加剧,其宏观效果便是金属的温度升高,表现为金属(导体)放出热量。

另一方面,正在被极化的电介质,其内位移电流应由性质截然不同的两部分组成,即在 $\boldsymbol{J}_d=\frac{\partial \boldsymbol{D}}{\partial t}=\varepsilon_0\frac{\partial \boldsymbol{E}}{\partial t}+\frac{\partial \boldsymbol{P}}{\partial t}$ 中,$\varepsilon_0\frac{\partial \boldsymbol{E}}{\partial t}$ 不涉及真实的电荷及其移动,只是激发磁场与 $\boldsymbol{J}_{传}$ 等效,而 $\frac{\partial \boldsymbol{P}}{\partial t}$ 等满足式(7-2-8),是真实的电荷移动,应具有与传导电流相似的性质。但由于电介质结构不同于金属导体的结构,束缚电流的热效应机制和结果应与电介质的类型和极化过程有关,而与金属导体中传导电流的热效应有相同点也有不同点。

**1. 无极分子介质极化束缚电流的热效应**

首先,就 $\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}=Nq\frac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}t}$ 本身来看,真实电荷的移动必然与介质原子、分子或离子晶格发生一定的碰撞或散射,从而发生能量交换而把其一部分定向动能传给介质原子、分子或离子晶格,则此时束缚电荷的能量交换完全等效于传导电流中自由电子与金属晶格的能量转移而使其原子、分子或晶格的振动能增加,其宏观上等效表现为温度升高而放出热量,这部分热效应应当具有与焦耳-楞次定律完全等效的形式,即

$$Q_p=\int_0^t i_p^2R'\mathrm{d}t=\int_0^t\left(\iint_S\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}\cdot\mathrm{d}\boldsymbol{s}\right)^2R'\mathrm{d}t=\int_0^t\left(\iint\boldsymbol{J}_P\cdot\mathrm{d}\boldsymbol{s}\right)^2R'\mathrm{d}t \tag{7-2-10a}$$

为简便计算,取均匀极化,且面元 $\mathrm{d}\boldsymbol{s}$ 与 $\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}$ 方向一致,并考虑式(7-2-8),则 $\iint_S\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}\cdot\mathrm{d}\boldsymbol{s}=\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}\cdot\boldsymbol{S}=Nq\frac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}t}\cdot\boldsymbol{S}=Nq\boldsymbol{v}(t)\cdot\boldsymbol{S}=Nqv(t)S$,其中 $\boldsymbol{v}(t)=\frac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}t}$ 是束缚电荷移动的速度,应是时间和外场频率函数,代入式(7-2-10a)得

$$Q_P=\int_0^t N^2q^2v^2(t)S^2R'\mathrm{d}t \tag{7-2-10b}$$

式(7-2-10a)和式(7-2-10b)中 $R'$ 是介质对束缚电荷电流的等效电阻,若引用焦耳-楞次定律的微分形式,则有

$$w=\frac{J_P^2}{\sigma'}=\left(\frac{\mathrm{d}P}{\mathrm{d}t}\right)^2/\sigma'=\frac{N^2q^2v^2(t)}{\sigma'} \tag{7-2-11}$$

式中 $\sigma'$ 是电介质对 $\boldsymbol{J}_P$ 的等效电导率。

按照式(7-2-10)的假定和讨论,这部分热效应等效于传导电流的热效应。无极分子的感生(电子)极化时,束缚电荷运动的热效应应主要是这种形式。

**2. 有极分子介质极化束缚电流的热效应**

有极分子介质化时,除了具有上述等效的焦耳热效应之外,由于偶极子的转向和离子偶极子的形成和转向,必然引起介质结构的变化,这种变化伴随有其他物理过程的发生,其中之一就是吸热和放热过程,但这种热效应不是焦耳热,不遵守该定律。

对有极分子介质在极化场较小时的极化,由统计物理可得出极化规律为

$$\boldsymbol{P}=\frac{NP_0^2}{3kT}\boldsymbol{E} \tag{7-2-12}$$

式中 $N$ 是单位体积内偶极子数，$\boldsymbol{P}_0$ 是每个偶极子的电偶矩的大小，$k$ 为玻耳兹曼常数，$T$ 为绝对温度，即其极化强度 $\boldsymbol{P}$ 与外场强 $\boldsymbol{E}$ 成正比而与温度成反比，说明有极分子介质极化的强度 $\boldsymbol{P}$ 要受到其本身热运动的影响，同时，$\boldsymbol{P}$ 的形成和取向也影响到介质的热运动性质，因为其结构发生了一定的变化。由式(7-2-12)得

$$\boldsymbol{J}_P=\frac{\partial \boldsymbol{P}}{\partial t}=\frac{NP_0}{3kT}\frac{\partial \boldsymbol{E}}{\partial t}-\frac{NP_0\boldsymbol{E}}{3kT^2}\frac{\partial T}{\partial t}+\frac{2NP_0q}{3kT}v\boldsymbol{E}=\boldsymbol{J}_E+\boldsymbol{J}_T+\boldsymbol{J}_P \tag{7-2-13}$$

式(7-2-13)表明：束缚电流不仅受外场变化的影响($\boldsymbol{J}_E$)也受到介质温度及其变化的影响($\boldsymbol{J}_T$)、($\boldsymbol{J}_P$)类似于式(7-2-8)及式(7-2-10)的讨论，不再赘述。由式(7-2-13)得

$$\boldsymbol{J}_T=-\frac{NP_0^2}{3kT^2}\frac{\mathrm{d}T}{\mathrm{d}t}\boldsymbol{E} \tag{7-2-14a}$$

式(7-2-14a)表明 $\boldsymbol{J}_T$ 与 $\boldsymbol{E}$ 方向相反，其大小关系为

$$J_T=\frac{NP_0^2E}{3kT^2}\frac{\mathrm{d}T}{\mathrm{d}t} \tag{7-2-14b}$$

即 $\boldsymbol{J}_T$ 引起电介质的温度变化为

$$\frac{\mathrm{d}T}{\mathrm{d}t}=\frac{3J_TkT^2}{NP_0^2E} \tag{7-2-15}$$

设介质的比热容为 $c$，单位体积的质量为 $m=\rho$，则单位时间内由于 $\boldsymbol{J}_T$ 部分引起该介质温度变化而放出$\left(\frac{\mathrm{d}T}{\mathrm{d}t}>0\right)$或吸收$\left(\frac{\mathrm{d}T}{\mathrm{d}t}<0\right)$的热量为

$$Q=cm\frac{\mathrm{d}T}{\mathrm{d}t}=\frac{3J_TkT^2\rho c}{NP_0^2E} \tag{7-2-16}$$

式(7-2-16)即是这种电介质的束缚电流热效应的另一种表现形式，这部分能量转移与焦耳热无关，是另一种机制。

### 7.2.4 结论

由式(7-2-9)

$$\nabla\times\boldsymbol{H}=\boldsymbol{J}_0+\boldsymbol{J}_\mathrm{d}=\boldsymbol{J}_0+\varepsilon_0\frac{\partial \boldsymbol{E}}{\partial t}+\frac{\partial \boldsymbol{P}}{\partial t} \tag{7-2-17}$$

涡旋磁场是由三部分电流贡献而构成，其一是自由电荷形成的传导电流 $\boldsymbol{J}_0$，$\boldsymbol{J}_0$ 的热效应遵守焦耳-楞次定律；其二是位移电流的第一部分即真空中的位移电流 $\varepsilon_0\frac{\partial \boldsymbol{E}}{\partial t}$，这部分位移电流没有热效应；其三是位移电流的第二部分即介质中束缚电荷电流$\frac{\partial \boldsymbol{P}}{\partial t}$，这部分位移电流应该具有传导电流的基本性质(见式(7-2-8))，其热效应一方面遵守焦耳-楞次定律(见式(7-2-10a)或式(7-2-11))，另一方面与电介质的结构类型和极化过程有关(见式(7-2-12)、式(7-2-13)及式(7-2-16))。

具体地说，在外场不太强时，无极分子介质极化产生的束缚电荷电流的热效应与传导电流的热效应机制基本一致，遵守焦耳-楞次定律；有极分子介质的转向极化时，一方面极化产生的束缚电荷电流也同无极分子介质中的一样，会产生焦耳-楞次热效应(有极分子也伴

有像无极分子那样的位移极性质的变化(如式(7-2-12)、(7-2-13)),相互影响的结果伴随有吸热或放热效应(如式(7-2-16)),这部分热效应显然不遵守焦耳-楞次定律。这是一个较为复杂的过程。

## 7.3　抗磁性的微观解释

某些物质在外磁场作用下,会呈现出抗磁性。

1822年,安培提出了物质磁性本质是分子电流。1911年,卢瑟福通过$\alpha$射线轰击各种原子的实验,从而确立了原子的核式结构的行星模型。卢瑟福的原子模型表明,原子是由带正电的原子核和带负电的电子组成。电子不仅绕核旋转,而且还自旋。电子的这些运动在原子、分子等微观粒子内部形成了"分子电流"。若这样一些分子电流定向排列起来,在宏观上就会显示N、S极,这就是物质磁性的基本来源。

根据物质的电结构理论,任何物质都是由分子或原子组成的。其中,除了各原子核和自旋运动外,它们所包含的每一个电子都参与两种运动:电子自旋以及电子绕原子核的轨道运动,这些运动都会形成微小的圆电流,都能产生磁效应。

磁矩是反映粒子本身这种特征的物理量,每种运动所对应的磁矩分别称为自旋磁矩和轨道磁矩。一个分子中的所有电子轨道磁矩和自旋磁矩以及核的自旋磁矩的矢量和,称为分子的固有磁矩,简称为分子磁矩,用$m_e$表示。每一个分子磁矩又可以用一个等效圆电流表示,称为分子电流。

在没有外磁场作用时,顺磁质中的每个分子磁矩不为零:$\boldsymbol{m}_e \neq 0$。但由于分子的热运动,分子磁矩排列的方向是杂乱无章的,因而对任一体积元$\Delta V$来说,所有分子磁矩的矢量和为零:$\sum \boldsymbol{m}_e = 0$。所以,对外不显磁性。有外磁场作用时,顺磁质的分子磁矩排列的方向将与外磁场方向趋于一致,从而顺磁质磁介质对外能显示出磁性,即$\sum \boldsymbol{m}_e \neq 0$。

抗磁质的每个分子磁矩为零:$\boldsymbol{m}_e = 0$。在没有外磁场作用时,所有分子磁矩的矢量和也为零:$\sum \boldsymbol{m}_e = 0$,宏观上不显磁性。然而,这类磁介质在外磁场的作用下,能产生抗磁性。

本节从三种不同的角度来解释磁介质的抗磁性,并对结果的等效性进行了讨论。

### 7.3.1　电子轨道运动在磁场中进动产生附加磁矩的解释

电子轨道磁矩与外磁场不平行(即夹角$0<\theta<\pi$),可用电子轨道运动在磁场中进动产生附加磁矩来解释,分两种情况。

**1. 电子轨道磁矩与外磁场的夹角:$0<\theta\leqslant\frac{\pi}{2}$**

如图7-3-1(a)所示,电子作轨道运动时,具有角动量

$$\boldsymbol{L} = mr^2\boldsymbol{\omega} \tag{7-3-1}$$

电子的轨道运动使它也具有磁矩

$$m_0 = IS = \frac{e\omega}{2\pi} \cdot \pi r^2 = -\frac{e}{2m}L \tag{7-3-2}$$

由于电子带负电,负号表示磁矩$\boldsymbol{m}_0$的方向与$\boldsymbol{L}$的方向相反。当加入外磁场时,电子的轨道磁矩要受力矩$\boldsymbol{M} = \boldsymbol{m}_0 \times \boldsymbol{B}$的作用,力矩的方向为垂直纸面向里。

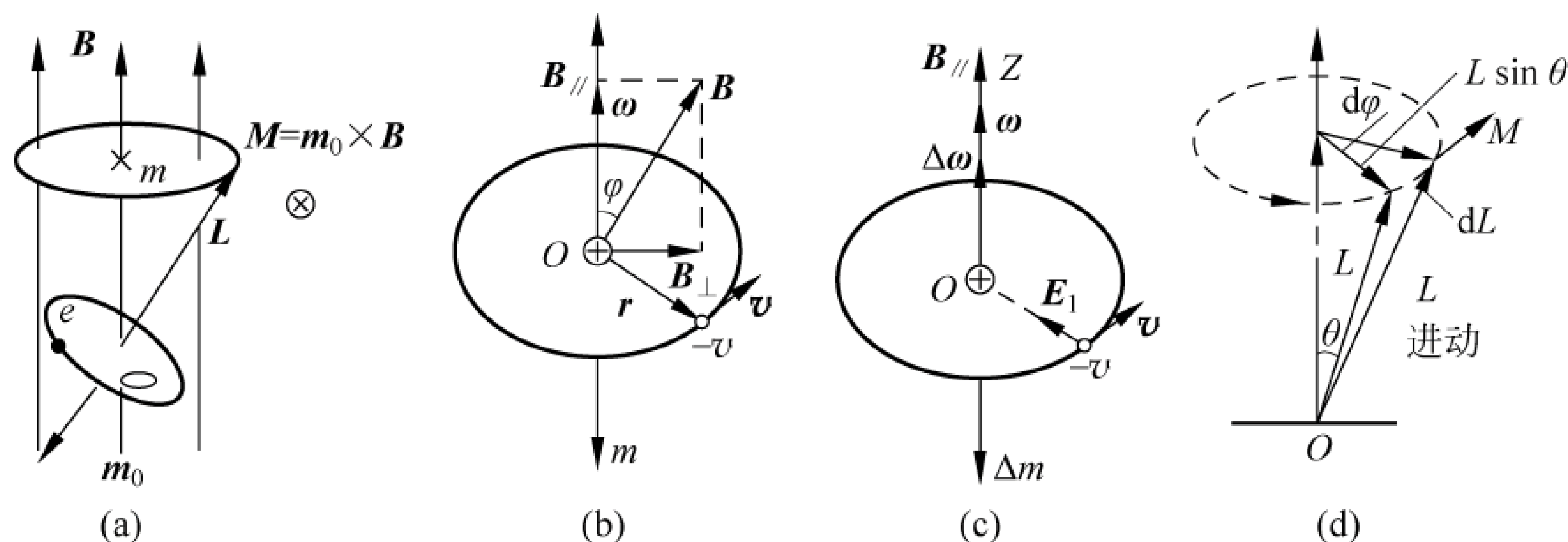



图 7-3-1　电子轨道磁矩与外磁场的关系

以电子的轨道磁矩 $\boldsymbol{m}$ 为 $z$ 轴，如图 7-3-1(b)所示，设电子的轨道角速度 $\omega$ 与外磁场 $\boldsymbol{B}$ 的夹角为 $\varphi$。则可将 $\boldsymbol{B}$ 矢量分解为：与 $\omega$ 平行的分量 $B_{/\!/}=B\cos\varphi$ 和与 $\omega$ 垂直的分量 $B_{\perp}=B\sin\varphi$。

为清楚起见，下面分两种情形来考虑每一分量各自对电子轨道运动的影响。

(1) 分量 $B_{/\!/}$产生的附加磁矩

如图 7-3-1(c)所示，电子除受核电场力 $F_e$ 作用外，同时还受磁场的洛伦兹力 $\boldsymbol{F}=-e\boldsymbol{v}\times\boldsymbol{B}_{/\!/}$的作用。由于 $B_{/\!/}$和 $\omega$ 之间方向相同，所以，洛伦兹力 $F_{\mathrm{m}}$ 的方向是指向轨道中心的，即与核电场力 $F_e$ 同向。电子运动方程就变为：$e\boldsymbol{E}+e\boldsymbol{v}\times\boldsymbol{B}_{/\!/}=m\omega^2\boldsymbol{r}$。由于轨道半径 $r$ 不变，则角速度的增量为 $\Delta\omega$ 时，其角速度将变成 $\omega=\omega_0+\Delta\omega$，相应的附加磁矩为

$$\Delta m=\Delta IS=\frac{e\Delta\omega}{2\pi}\cdot\pi r^2=-\frac{er^2}{2}\Delta\omega \tag{7-3-3}$$

一般说来，只要外磁场 $\boldsymbol{B}$ 不太大的话，角速度增量的大小 $\Delta\omega\ll\omega_0$，所以产生的附加磁矩往往可以忽略不计。

(2) 分量 $B_{\perp}$产生的附加磁矩

如图 7-3-1(d)所示，电子在磁场 $B_{\perp}$的作用下受到的磁力矩为

$$\boldsymbol{M}=\boldsymbol{m}\times\boldsymbol{B}_{\perp}=-\frac{e}{2m}\boldsymbol{L}\times\boldsymbol{B}_{\perp}=\frac{e}{2m}\boldsymbol{B}_{\perp}\times\boldsymbol{L} \tag{7-3-4}$$

又根据动量矩定理 $\boldsymbol{M}=\dfrac{\mathrm{d}\boldsymbol{L}}{\mathrm{d}t}$可知，磁力矩 $\boldsymbol{M}$ 在 $\mathrm{d}t$ 时间内所引起的角动量的增量 $\mathrm{d}\boldsymbol{L}$ 的方向与 $\boldsymbol{M}$ 的方向一致，因而与 $\boldsymbol{L}$ 垂直。如图 7-3-1(d)所示，在 $\mathrm{d}t$ 时间内，角动量由 $\boldsymbol{L}$ 变为 $\boldsymbol{L}+\mathrm{d}\boldsymbol{L}$，其指向则相应于绕 $\boldsymbol{B}$ 为轴转过一角度 $\mathrm{d}\varphi$，而角动量 $\boldsymbol{L}+\mathrm{d}\boldsymbol{L}$ 与 $\boldsymbol{B}$ 轴的夹角仍为 $\varphi$，和原先 $\boldsymbol{L}$ 与 $\boldsymbol{B}$ 轴之间的夹角相等。当角动量矢量 $\boldsymbol{L}$ 转到新位置 $\boldsymbol{L}+\mathrm{d}\boldsymbol{L}$ 时，因 $B$ 的大小、角动量矢量的大小及两者之间的夹角 $\varphi$ 大小均不变，故由式(7-3-4)可知，磁力矩的大小 $|M|$ 仍保持不变，只是磁力矩 $\boldsymbol{M}$ 的方向随着角动量矢量 $\boldsymbol{L}$ 的方向改变成 $\boldsymbol{L}+\mathrm{d}\boldsymbol{L}$ 而发生了改变。由图 7-3-1(d)可见，这时磁力矩 $\boldsymbol{M}$ 的方向仍与新位置的角动量矢量 $\boldsymbol{L}+\mathrm{d}\boldsymbol{L}$ 垂直。所以，磁力矩 $\boldsymbol{M}$ 继续使角动量矢量 $\boldsymbol{L}$ 绕 $B$ 轴旋转。

由此可知，在外磁场分量 $B_{\perp}$的作用下，绕核轨道运动的电子的角动量矢量 $\boldsymbol{L}$ 的大小始终不变，只是角动量矢量的方向在改变，即绕 $\boldsymbol{B}$ 为轴沿一圆锥面旋转。因此，角动量 $\boldsymbol{L}$ 在外磁场 $\boldsymbol{B}$ 方向上的投影 $L\cos\varphi$ 也保持大小不变。这是因为外磁场 $\boldsymbol{B}$ 在垂直于角动量 $\boldsymbol{L}$ 方向(在不断变化着)上的投影分量 $B_{\perp}=B\sin\varphi$ 的大小保持不变，这样一种运动就是所谓的进动。

整个来看，在外磁场 $\boldsymbol{B}$ 中电子的运动可视作电子除了继续在轨道平面内仍绕核快速转动外，同时轨道平面还不断绕着 $\boldsymbol{B}$ 方向为轴作进动。这样进动的角速度称拉莫尔(Larmor Joseph，1857—1942)角频率，它表示加外磁场 $B_{\perp}$ 时，电子绕核转动角速度 $\omega$ 方向对时间的变化率。而电子进动的方向由 $\boldsymbol{M}$ 决定，如图 7-3-1(d)所示。

设在 $\mathrm{d}t$ 时间内，角动量的变化为 $\mathrm{d}\boldsymbol{L}$，由图 7-3-1(d)可得 $\mathrm{d}L=L\sin\theta\cdot\mathrm{d}\varphi$。

根据动量矩定理 $\boldsymbol{M}=\frac{\mathrm{d}\boldsymbol{L}}{\mathrm{d}t}$，而 $M=\frac{e}{2m}BL\sin\theta$。则电子角动量 $\boldsymbol{L}$ 因进动转过的角度为 $\mathrm{d}\varphi=\frac{\mathrm{d}L}{L\sin\theta}=\frac{M\mathrm{d}t}{L\sin\theta}=\frac{\frac{e}{2m}BL\sin\theta}{L\sin\theta}\mathrm{d}t=\frac{e}{2m}B\mathrm{d}t$ 则电子进动的角速度为 $\omega_L=\frac{\mathrm{d}\varphi}{\mathrm{d}t}=\frac{e}{2m}B$。与这一进动相应，除了原有的轨道磁矩 $m_0$ 外，还增加了一个附加磁矩 $\Delta m$，该磁矩的方向与外磁场的方向相反，大小为

$$\Delta m=I'S'=\frac{e\omega_L}{2\pi}\cdot\pi R^2=\frac{e\frac{e}{2m}B}{2\pi}\cdot\pi R^2=\frac{e^2}{4m}BR^2 \tag{7-3-5}$$

其中，$R$ 为电子进动时的轨道半径。

由于电子进动而产生的附加磁矩 $\Delta m$ 与外磁场 $\boldsymbol{B}$ 方向相反，即产生了抗磁性。

**2. 电子轨道磁矩与外磁场的夹角：$\frac{\pi}{2}\leqslant\theta<\pi$**

如图 7-3-2 所示，与图 7-3-1 作相反方向轨道运动的电子，外磁场将对电子的轨道运动产生相反的力矩 $\boldsymbol{M}$，使角动量 $\boldsymbol{L}$ 沿与 $\boldsymbol{B}$ 平行的轴进动，与这一进动相应所产生的附加磁矩 $\Delta\boldsymbol{m}$ 也和外磁场 $\boldsymbol{B}$ 的方向相反，即产生抗磁性，其大小与图 7-3-1 中的情形相同。

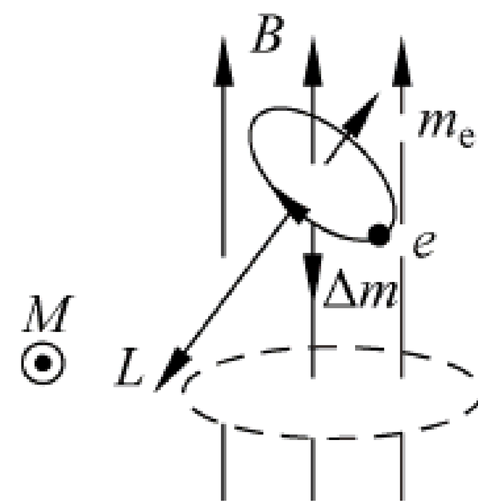


图 7-3-2　电子轨道磁矩与外磁场的关系

## 7.3.2 电磁感应附加磁矩的解释

上面讨论的是电子轨道角动量与外磁场夹角不平行的情形，这是因为当电子轨道角动量与外磁场平行和反平行时，$\boldsymbol{M}=\boldsymbol{m}_0\times\boldsymbol{B}$ 进动是不存在的。而这种情况可用电磁感应附加磁矩来解释。

这种解释是用电子轨道运动在磁场中感生电场来说明附加磁场产生过程。由于电子沿着圆形轨道运动时，可以等效为圆形电流，从而具有相应的轨道磁矩。在外磁场加入磁介质的过程中，如果该磁场是变化的，就会产生感生电场，感生电场会使磁介质中的电子在轨道上的运动速率改变，从而引发附加的等效电流，即感应电流。感应电流的磁矩方向总是与外加磁场方向相反，这就是产生抗磁性的原因。

设质量为 $m$，电量为 $e$ 的电子以速度 $v$ 沿半径为 $r$ 的轨道作圆周运动，因为电子每秒转 $\frac{v}{2\pi r}$ 圈，所以等效电流为 $I=\frac{ev}{2\pi r}$。下面分两种情形讨论。

**1. 电子轨道磁矩与外磁场平行**

如图 7-3-3(a)所示，按经典模型，一电子沿半径为 $r$ 的圆形轨道运动速率为 $v_0$，则电子的轨道运动的磁矩为 $m_0=IS=-\frac{ev_0}{2\pi r}\cdot\pi r^2=-\frac{ev_0r}{2}$(方向与电子的轨道角动量相反)。

如果垂直于轨道平面加一磁场 $B$，如图 7-3-3(b)所示，其方向向上，则在外加磁场由 0

增大到 $B$ 的过程中，在电子运动区域将产生感生电场 $E$。由法拉第定律得 $\varepsilon=\int_l \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=-\dfrac{\mathrm{d}\Phi}{\mathrm{d}t}$，代入 $\dfrac{\mathrm{d}\Phi}{\mathrm{d}t}=\dfrac{\mathrm{d}}{\mathrm{d}t}\int_s \boldsymbol{B}\,\mathrm{d}s=\pi r^2\dfrac{\mathrm{d}\boldsymbol{B}}{\mathrm{d}t}$，可得：$E2\pi r=-\pi r^2\dfrac{\mathrm{d}B}{\mathrm{d}t}$，所以

$$\boldsymbol{E}=-\frac{r}{2}\frac{\mathrm{d}\boldsymbol{B}}{\mathrm{d}t} \tag{7-3-6}$$

初始状态 (a)　　最终状态 (b)

图 7-3-3　电子轨道磁矩与外磁场平行

在此电场作用下，电子轨道运动的轨道半径不变，电子将沿轨道加速，加速度为

$$a=\frac{F}{m}=-\frac{eE}{m}=\frac{er}{2m}\frac{\mathrm{d}B}{\mathrm{d}t} \tag{7-3-7}$$

在轨道半径 $r$ 不变的情况下，电子速率的增加值为

$$\Delta v=\int a\,\mathrm{d}t=\int_0^B \frac{er}{2m}\mathrm{d}B=\frac{er}{2m}B \tag{7-3-8}$$

则与之相应的电子轨道磁矩增量，即附加磁矩大小为

$$\Delta m=-\frac{e\Delta vr}{2}=-\frac{e^2r^2}{4m}B \tag{7-3-9}$$

其方向由速度的增量的方向判断，如图 7-3-3(b)所示和外磁场的方向相反。

**2. 电子轨道磁矩与外磁场反平行**

如果电子轨道磁矩与外磁场反平行(假设外磁场方向竖直向上)，则无外磁场时，电子的磁矩方向将向上。如图 7-3-4(a)所示。

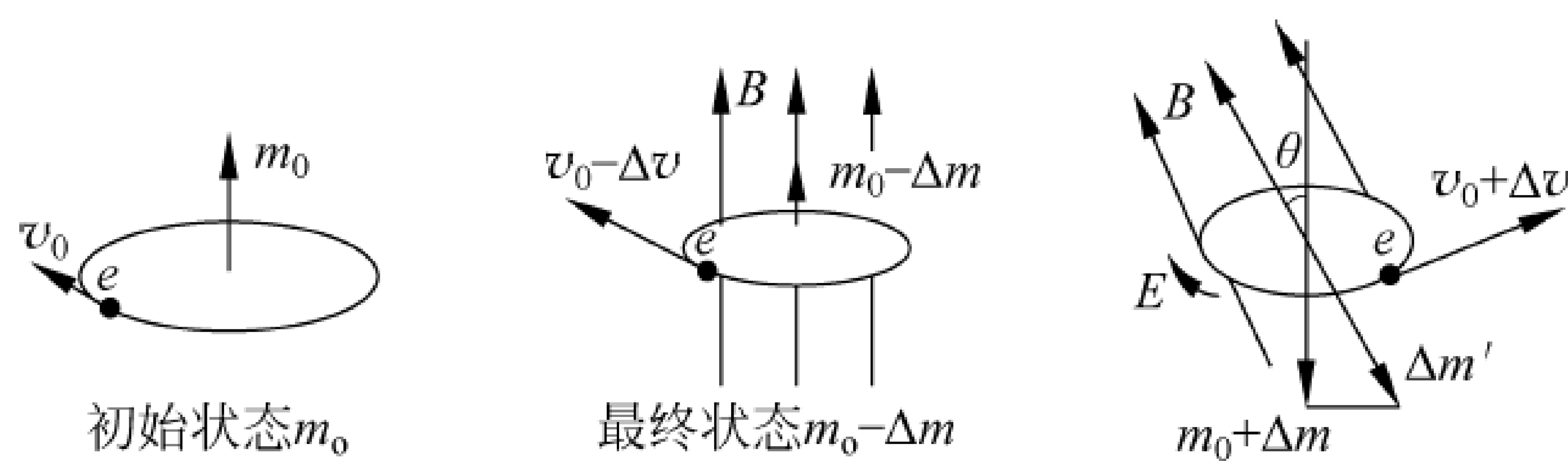



图 7-3-4　电子轨道磁矩与外磁场反平行

在加外磁场的过程中，如图 7-3-4(b)所示，感生电场将使电子减速 $\Delta v$，与之相应电子将减少与外磁场方向相反的附加磁矩 $\Delta m$，其大小与图 7-3-1 中的情形相同。

上述解释可推广到电子轨道磁矩方向与外磁场有任意角度时的情况。

由图 7-3-4 可看出，对于外磁场沿任意方向时，抗磁质的附加磁矩总是与外磁场方向相反的，同样呈现抗磁性。此时产生的附加磁矩大小为

$$\Delta m=-\frac{e\Delta vr}{2}=-\frac{e^2r^2}{4m}B\cos\theta \tag{7-3-10}$$

由此看到，用电磁感应产生附加磁矩能较好地解释电子轨道磁矩任意取向的情形。

### 7.3.3 经典统计方法的解释

从原子在外磁场中能量的改变出发，对于电子轨道角动量不平行于外磁场方向的任意情况，利用经典统计原理，可得到与前者完全一致的结果。由经典统计解释物质的抗磁性，可使我们对附加磁矩恒与外磁场方向相反的理解更加准确清楚、全面。

为了简单起见，我们仅讨论原子核外只有一个电子的介质的磁化(对多电子原子，只须

把下面的角动量理解为所有电子角动量之和，最终结果乘以原子序数 $z$，结果中的 $r$ 换为平均值 $r$ 即可）。

该介质原子的固有磁矩为 $\boldsymbol{m}=-\frac{e}{2m_e}\boldsymbol{L}=-\frac{e}{2m_e}\boldsymbol{r}\times(\boldsymbol{\omega}\times\boldsymbol{r})$。当加上外磁场 $B$ 后，电子除了绕核作轨道运动外，还要产生一种附加的进动，即拉莫尔进动，其进动角速度为

$$\boldsymbol{\omega}_L=\frac{e}{2m_e}\boldsymbol{B} \tag{7-3-11}$$

由于拉莫尔进动，原子在外磁场中的磁矩变为

$$\boldsymbol{m}=-\frac{e}{2m_e}\boldsymbol{r}\times(\boldsymbol{\omega}\times\boldsymbol{r})=\frac{e}{2m_e}\boldsymbol{L}-\frac{e_2}{4m_e}[r^2\boldsymbol{B}-(\boldsymbol{r}\cdot\boldsymbol{B})\boldsymbol{r}] \tag{7-3-12}$$

式中第二项正是电子进动产生的附加磁矩。

物质的抗磁性是磁场对电子轨道运动的作用结果，应该发生在任何原子或分子中，因此是普遍存在的。但只有在所有电子的轨道角动量完全抵消时，抗磁性才显示出来；如果不是这样，较强的顺磁性会掩盖抗磁性。所以，一旦介质出现抗磁性时，则式中第一项为零，这时介质中原子的磁矩就等于附加磁矩，即 $\Delta\boldsymbol{m}=-\frac{e_2}{4m_e}[r^2\boldsymbol{B}-(\boldsymbol{r}\cdot\boldsymbol{B})\boldsymbol{r}]$。

根据经典电磁理论，具有上式磁矩 $m$ 的原子在外磁场 $B$ 中的能量改变为

$$\Delta E=\Delta\boldsymbol{m}\cdot\boldsymbol{B}=\frac{e_2}{4m_e}r^2B^2\sin^2\theta \tag{7-3-13}$$

我们把磁介质看作是与外界这个大热库接触的系统，则原子按能量的分布为正态分布，即原子处于能量为 $E$ 的状态的概率：$\rho\propto e^{-\frac{E}{kT}}$。当加上外磁场并达到平衡后，原子按能量的分布就完全决定于附加能量 $\Delta E$，即 $\rho=ce^{-\frac{\Delta E}{kT}}$。代入 $\Delta E$ 得

$$\rho=c\exp\left(-\frac{e^2}{4m_ekT}r^2B^2\sin^2\theta\right) \tag{7-3-14}$$

上式即为在磁场 $B$ 中，原子产生附加磁矩的概率。

设 $N$ 为单位体积内的磁原子数，则上式两边乘以 $N$ 后，即为单位体积内产生附加磁矩的原子数 $n(\theta)$，即 $n(\theta)=n_0\exp\left(-\frac{e^2}{4m_ekT}r^2B^2\sin^2\theta\right)$，其中 $n_0$ 为待定常数。

在通常温度和磁场条件下，上式指数部分远远小于 1，所以按指数展开，取一级近似后上式化为：$n(\theta)=n_0\left(1-\frac{e^2}{4m_ekT}r^2B^2\sin^2\theta\right)$。由于 $\int_0^\pi n(\theta)2\pi\sin\theta\,d\theta=N$，故可定出常数为

$$n_0=N/\left(4\pi-\frac{2\pi e^2r^2B^2}{3m_ekT}\right) \tag{7-3-15}$$

根据磁化强度的定义，磁化强度 $\boldsymbol{M}$ 等于单位体积内所有原子磁矩的矢量和。由磁介质中磁原子分布关于外磁场方向的均匀对称性可知，介质中任何一小体元内各原子附加磁矩的矢量和只可能在外磁场 $\boldsymbol{B}$ 方向上，而附加磁矩的其余分量相互抵消。另外，从时间平均来看，附加磁矩在垂直于外磁场方向的分量的平均效果完全抵消。因此，总磁化强度就等于单位体积内各原子附加磁矩的外磁场方向分量之和，即

$$M=\int_0^\pi n(\theta)\left(-\frac{e^2}{4m_e}r^2B^2\sin^2\theta\right)2\pi\sin\theta\,d\theta \tag{7-3-16}$$

代入 $n(\theta)$ 和 $n_0$，可得 $M=-\frac{Ne^2r^2B^2}{6m_e}\left[\frac{2-2e^2r^2B^2/(5m_e kT)}{2-e^2r^2B^2/(3m_e kT)}\right]$。在通常低温条件(如 $T<10^{-3}$K)下，上式中括号项非常接近于1，于是有 $M=-\frac{Ne^2r^2B^2}{6m_e}$。利用关系式 $M=\chi_m H$ 和 $B=\mu H$，对于弱磁性介质，$\frac{N\mu_0 e^2r^2}{6m_e}\ll 1$，可得磁化率近似为

$$\chi_m=-\frac{N\mu_0 e^2 r^2}{6m_e} \tag{7-3-17}$$

因为磁化率为负，所以介质表现出抗磁性，磁化率中不包括温度 $T$，说明抗磁性与温度无关。该结果与上述其他方法所得结果完全一致。

### 7.3.4 结语

一个分子内的所有电子或原子核都会产生与外磁场方向相反的附加磁矩。

顺磁质的分子具有固有磁矩 $\boldsymbol{m}_0$，在无外磁场时，由于分子运动的不规则性，整个顺磁质分子磁矩总和为零，不显磁性。当加入外磁场 $\boldsymbol{B}$ 后，在外磁场作用下，也会产生拉莫尔进动，产生一个反向的附加磁矩，即表现为抗磁性。这些附加磁矩的矢量和 $\sum\Delta\boldsymbol{m}$ 就是磁介质在外磁场中所产生的感生磁矩。用磁介质的磁化强度 $\boldsymbol{M}=\frac{\sum\Delta\boldsymbol{m}}{\Delta V}$ 表征磁介质的宏观抗磁性。但是由于分子存在固有磁矩，在外磁场的作用下，每个分子的固有磁矩均要趋向外磁场方向，这是顺磁效应。对于顺磁质，抗磁效应和顺磁效应都是存在的。但是抗磁效应比顺磁效应弱得多，顺磁质的抗磁效应被顺磁效应所掩盖，因而表现出顺磁性。由于顺磁效应与温度有关，而抗磁效应却与温度无关，所以在实验上不难将这两者区分开来。

而抗磁质的所有附加磁矩 $\sum\Delta\boldsymbol{m}$ 方向总是与外磁场 $\boldsymbol{B}$ 方向相反，即 $\boldsymbol{M}$ 的方向与外磁场 $\boldsymbol{B}$ 方向相反，因此，宏观上便显示出抗磁性。

## 7.4 电场与磁场的相对性

根据相对论的观点，同一电磁场在不同惯性系中所观察到的电场和磁场是不同的，这就是说电场与磁场具有相对性。现从相对论的角度通过一个实例对此作一分析。

设有一带正电 $q$ 的粒子正以速度 $u$ 平行于一载流长直导线运动。它离导线轴线的距离为 $r$，如图 7-4-1 所示。现分别从固定于导线上的参考系 $k$ 和固定在带电粒子上的参考系 $k'$ 来分析这个带电粒子的受力情况。

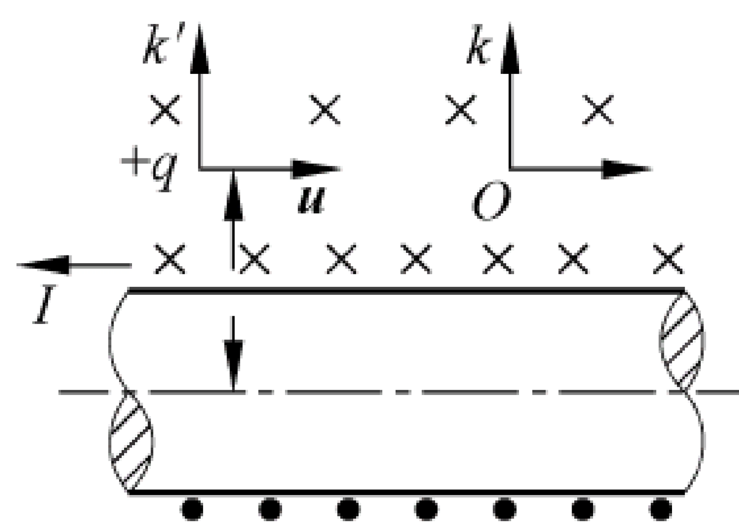


图 7-4-1 正电荷 $q$ 在不同参考系的受力不同

### 7.4.1 从静参考系 $k$ 来分析

在参考系 $k$，导线是静止的，导线中的自由电子以平均速度 $\boldsymbol{v}$ 向右作定向运动，形成了向左的稳恒电流 $I$。我们假定导线均匀且远离电源，因此，当通过稳恒电流时导线上不存在净电荷。如图 7-4-1 设导线中的负电荷体密度为 $\rho_-$，正电荷的体密度为 $\rho_+$，且 $|\rho_+|=|\rho_-|=\rho$，所以，在导线外不存在电场。设导线外一正电荷 $q$ 的速度为 $\boldsymbol{u}$ 也向右运动，作用在该运动正电荷 $q$ 上的力就只有由载流导线周围的磁场作用在 $q$ 上的洛伦兹力 $\boldsymbol{F}$，

其方向垂直背离导线向上，因而该正电荷将偏离导线向右上方运动，实验结果亦如此。由安培环路定律算得带电粒子所在处的磁场大小为：$B=\frac{u_0 I}{2\pi r}=\frac{u_0}{2\pi}\frac{\rho S v}{r}$（其中 $S$ 为导线的横截面积，$I=JS=\rho S v$），洛伦兹力 $\boldsymbol{F}=q\boldsymbol{u}\times\boldsymbol{B}$，将上式代入，又考虑到 $\boldsymbol{u}\perp\boldsymbol{B}$，得 $F=\frac{u_0}{2\pi}\frac{q\rho S u v}{r}$。又因 $u_0\varepsilon_0=\frac{1}{c^2}$，所以，又可以写成 $F=\frac{q\rho S}{2\pi\varepsilon_0 r}\frac{vu}{c^2}$，为简单计算，讨论 $u=v$ 的情况，这时

$$F=\frac{q\rho_- S}{2\pi\varepsilon_0 r}\frac{u^2}{c^2} \tag{7-4-1}$$

## 7.4.2　从动参考系 $k'$ 来分析

设一动参考系 $k'$ 随自由电子以平均速度 $\boldsymbol{v}$ 向右作定向运动，对 $k'$ 来说自由电子无定向漂移运动。因此，不会受到洛伦兹力的作用。对载流导线本身来分析，导线中有正离子和自由电子，在 $k$ 系中静止的正离子的体密度 $\rho_+$ 和以速度 $\boldsymbol{v}$ 运动的自由电子的体密度 $\rho_-$ 正好相等。从 $k'$ 系来看，这时正离子有一个向左的速度 $u$（为讨论方便，现令其大小相等：$u=v$）。由于导线上的总电量是不随参考系而变的，根据相对论，运动物体在运动方向上的长度将缩短，即正离子的间距将变小。因而，在 $k'$ 系中正电荷的体密度 $\rho'_+$ 将变大（$\rho'_+>\rho_+$）。同样，在 $k$ 系中，自由电子是静止的（在 $u=v$ 的情况），根据相对论，电子的间距要变大，在总电量不变下就导致 $k'$ 系负电荷密度 $\rho'_-$ 变小（$\rho'_-<|\rho_-|$）。因为 $\rho_+=|\rho_-|$，结果因参考系的变化导致 $\rho'_+>\rho_-$，导线在 $k'$ 系显示带正电，从而在导线周围出现了静电场，电荷在 $k'$ 系中所受到的向上的力就是静电力。下面对这个定性的解释作一些定量计算。

### 1. 在 $k'$ 系中导线上出现的净电荷密度

设在 $k$ 系中导线长度为 $L$，截面积为 $S$，导线中对 $k$ 系静止的某种带电粒子的电荷密度为 $\rho$，则它的总电量 $Q=\rho S_1$，在 $k'$ 系中该导线以 $\boldsymbol{v}$ 沿长度方向向右运动，根据相对论效应，其长度变为 $L'=L\sqrt{1-v^2/c^2}$，而在垂直于运动速度方向的长度不变，所以 $S'=S$，这时的总电量应为 $Q'=\rho' LS$（其中 $\rho'_+$ 为带电粒子在 $k'$ 中的电荷密度），据电量不随参考系改变，即 $Q=Q'$，所以，$\rho' L'=\rho L$，代入长度相对论关系可得

$$\rho'=\rho/\sqrt{1-v^2/c^2} \tag{7-4-2}$$

这是由相对论得出的电荷密度变化式，把式(7-4-2)应用到导线中的正离子，得

$$\rho'_+=\rho_+/\sqrt{1-v^2/c^2} \tag{7-4-3}$$

应用到导线中的自由电子，得

$$\rho_-=\rho'_-/\sqrt{1-v^2/c^2} \tag{7-4-4}$$

（注意，自由电子在 $k'$ 系中是静止的，在 $k$ 系中是速度 $\boldsymbol{v}$ 运动的），式(7-4-4)可变为

$$\rho'_-=\rho_-\sqrt{1-v^2/c^2} \tag{7-4-5}$$

因此，在 $k'$ 系中该段导线上出现的净电荷密度为 $\rho'=\rho'_+-\rho'_-=\frac{\rho_+}{\sqrt{1-v^2/c^2}}-\rho_-\sqrt{1-v^2/c^2}$。考虑到在 $k$ 系中导线为中性，即 $\rho_+=\rho_-$，所以，式(7-4-5)变为

$$\rho'=\frac{\rho_+}{\sqrt{1-v^2/c^2}}\cdot\frac{v^2}{c^2} \tag{7-4-6}$$

这就是在 $k'$ 系中有相对论效应时运动导线上出现的净电荷密度(正电荷),显然 $\rho'\neq 0$。

**2. 计算在 $k'$ 系中电荷 $q$ 所受的静电场力**

根据以上分析,在 $k'$ 系,运动导线带正电,因而就在其周围激发静电场,由高斯定理可求得均匀无限长直带电导线周围的场强大小为 $E=\dfrac{\lambda}{2\pi\varepsilon_0 r}$,其中 $\lambda$ 为导线上电荷的线密度,因而 $\lambda=\rho' S$,得导线外一运动正电荷 $q$ 所受到的向上的静电力大小为

$$F'=qE=q\frac{\rho' S}{2\pi\varepsilon_0 r} \tag{7-4-7}$$

把式(7-4-6)代入式(7-4-7)得

$$F'=qE=q\frac{S}{2\pi\varepsilon_0 r}\frac{\rho_+}{\sqrt{1-v^2/c^2}}\cdot\frac{v^2}{c^2} \tag{7-4-8}$$

从定量计算可见,导线外一运动正电荷 $q$ 在 $k$ 系中只受到洛伦兹力 $F$,而在 $k'$ 系中只受到库仑力 $F'$,两者方向相同,大小不同,且力的性质不同。当 $v\ll c$ 时,$\sqrt{1-v^2/c^2}\approx 1$,这时式(7-4-1)与式(7-4-8)相等,且 $\boldsymbol{F}=\boldsymbol{F}'$,这就是说在低速情况下,在不同惯性系中看到电荷 $q$ 的受力大小和方向是一致的。

通过这个实例的分析,可看到电磁场是相对的,在 $k$ 系中看来,载流长直导线周围只有磁场(因 $\rho_+=\rho_-$ 无电场),而在 $k'$ 系中看来该导线周围只有电场,应该认为,电磁场是一个客观事物,它独立于观察者而存在。无论从 $k$ 系还是 $k'$ 系来考察,只是从不同的参考系看到的"侧面"不同罢了,因此电和磁,电场和磁场,电场力和磁场力都是同一个东西的两个不同的方面,它们都有相对性,这种区分是随参考系的选择不同而变化的。客观上存在的电磁现象,电磁场和电磁力是一个整体。

关于电场和磁场的相对性,有兴趣的读者可以参阅张三慧主编,由清华大学出版社出版的《大学物理学·电磁学》第 2 章的内容。

## 7.5 电磁学中的场量与路量的比较

人类关于力的概念最初是由观察物体在直接接触时所发生的相互作用而产生的。这种"接触作用力"的观念曾经处于统治地位,成了人们的一般常识。虽然重力现象和磁力现象早为人所知,但是人们并没有想到这些现象有什么"非接触"作用的性质。

### 7.5.1 什么是场

**1. 超距作用对人们的影响**

牛顿理论清楚地告诉人们,万物之间存在着一种吸引力,这个力的方向在物体之间的连线上,大小与物体间的距离的平方成反比;在某一时刻作用于物体上的力只决定于同一时刻其他物体的位置,而与物体的速度和加速度无关;物体之间传递引力信息不需要时间,这就是所谓牛顿的超距作用思想。由此可以看出,牛顿的超距思想有三个特征:第一,瞬时性;第二,沿物体中心的连线;第三,不受空间介质的影响。

在"接触作用"和"超距作用"这两种完全相反的概念面前,具有不同哲学思想的人,态度是不同的。牛顿的后继者们在某些方面甚至比牛顿走得更远。他们企图把所有看来是"接触作用"的力都纳入超距作用的范畴。在牛顿的超距作用理论的影响下建立的电磁学理论有库仑定律、安培环路定律等。

（1）库仑定律

牛顿在《自然哲学的数学原理》中证明过，如果引力是随距离平方的增大而减少的话，中空的球体对球体内部的物体就没有作用力。1766年，美国物理学家富兰克林在实验中发现：放在金属杯中的软木塞，完全不受金属杯电性的影响。同年英国化学家普利斯特利(J. Joseph Priestley，1733—1804)也发现：当中空的金属器带电时，除开口的地方外，它对空腔内的电荷无任何作用力。于是，他大胆假设：电荷间的作用力也类似物体间的引力。这样他将牛顿的引力思想移植到了电磁学中。当时他本人并未证明，而为后来科学家的研究指示了一个方向。

1785年，法国科学家库仑利用扭秤实验验证了反比定律——库仑定律。此定律既继承了牛顿引力定律的形式，也继承了牛顿的超距作用观念。同时也说明了引力和电力是同类力，是统一的。

（2）安培定律

1820年，法国的物理学家阿喇果(Arago，Dominique Francois Jean，1786—1853)在瑞士听到奥斯特发现了电流磁效应的消息，立即向法国科学院报告了这一情况。安培听后，重复了奥斯特的实验，并加以发展。经过一系列经典研究和简单实验，安培为“磁”的本质是电提供了确切的依据。在研究电流的相互作用时，安培像牛顿那样将电流看成无数个电流元的集合，仿照牛顿力学，将电磁力转化为电流元之间的吸引力与排斥力，然后，他总结出了两个电流元之间的相互作用力，即平方反比的作用力形式。此理论保证了将电磁现象纳入了牛顿力学范畴，说明了电力、磁力、引力是同类力，是统一的。

**2. 近距理论的起源**

牛顿的后继者中另外一些人则选择了一条不同的道路，他们把看来像“超距作用”的力看作是靠充满空间的一种介质传递的，这样就把超距作用归并到接触作用的概念之中了。这种观念导致了“以太”假说的出现。

笛卡儿于1644年最先将以太引入科学，并提出了引力的“以太旋涡说”。他认为，物质是连续的，在这样的物质世界中，不存在任何超距作用，只有物体相互接触才能产生运动，物体之间的所有作用力都必须通过某种中间媒介物来传递。因此，虚空是不存在的，整个空间充满着一种看不见的本原物质，这就是以太。他认为，以太处于不停的激烈运动之中，各部分相互作用形成许多不同大小、速度和密度的旋涡。在太阳周围有一个巨大的旋涡，它的回旋带动了地球和其他行星。每个行星都处在产生通常的重力现象的另一个旋涡中心。更小的旋涡使物体各部分之间产生内聚力。

**3. 奥斯特的发现**

1820年4月，丹麦物理学家奥斯特安排了这样一个实验，把导线与磁针都沿磁子午线方向平行放置，接通电源后，小磁针向垂直于导线的方向偏转过去。他在1820年7月21日以《关于磁体周围电冲突的实验》为题，记载到：“在自由悬挂的磁针上方，由北向南流动的伽伐尼电，把磁针的北端推向东，而在相同的方向上，在磁针下面流过的伽伐尼电，把磁针的北端推向西。”奥斯特在实验中还发现，使用不同种类的金属导体，通电后都能使磁针偏转，只有程度上的差别，但是将玻璃、木头等非磁性物质置于导线和磁针之间，效应不受影响。奥斯特把导体中及其周围空间发生的这种效应称为“电冲突”(electric conflict)，并得出结论：电冲突只能对磁性粒子起作用。奥斯特的发现与牛顿的超距作用理论是矛盾的。因为超距理论认为力只能是沿作用于物体连线上的吸引或排斥，但是奥斯特发现的是一种“旋转力”。

**4. 法拉第关于场的思想**

法拉第根据奥斯特的“电冲突”在载流导线周围呈圆形分布的思想，于1821年9月，在一个盛着水银的容器中间铅直地固定一根磁棒，在它上方悬起一根硬导线，导线斜着浸入水银液内，如图7-5-1所示。然后将导线接通电流，结果发现导线下端绕着磁棒缓慢地作圆周运动。这就是电磁转动效应的实验。电磁转动的研究工作，使法拉第产生一种直觉，他意识到环绕着载流导线有一个圆形作用“线”；他设想这是当电流通过导线时在导线内的分子和周围的媒质中引起的某种“紧张”状态。

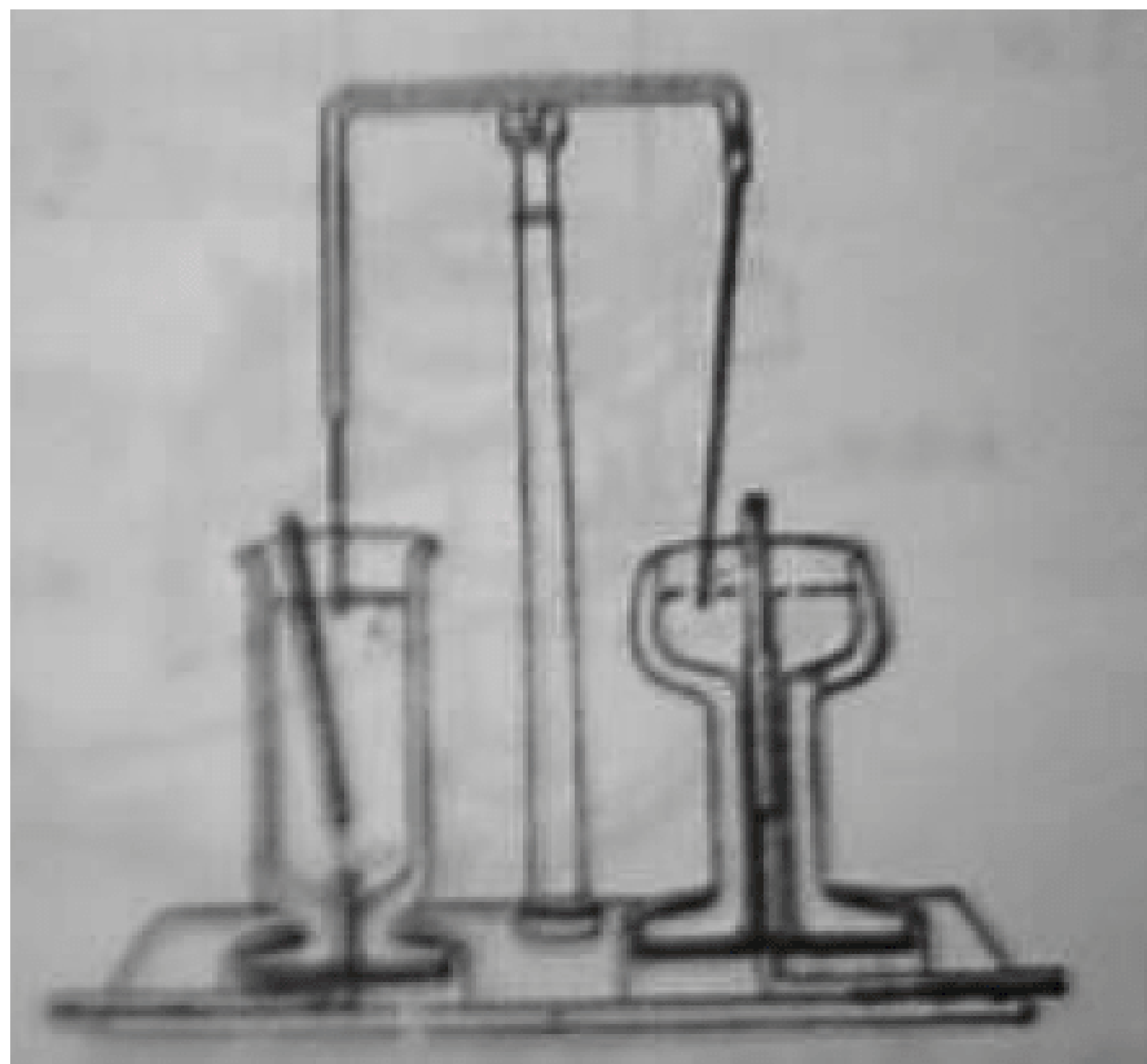


图7-5-1 法拉第的电磁转动实验装置

1831年10月17日，他在一个圆纸筒上绕了多层线圈，将一个圆柱形的磁棒迅速地插入螺线圈中，这时线圈所连的电流计的指针动了，抽出磁棒时，指针又动了，但偏转方向相反。在1831年11月24日，法拉第向皇家学会提交的一个报告中，把这种现象定名为“电磁感应”。法拉第基于他所发现的“电磁感应”现象，提出了“电致紧张状态”和“磁感线”两个概念。法拉第认为，“电致紧张状态”是由电流或磁体产生的存在于物体和空间中的张力状态，这种状态的产生、变化和消失，都会使处于这种状态中的导体感应出电流来。“磁感线”是法拉第为了对物体与螺线管之间的相对运动产生感应电流的现象进行解释而提出的一个概念。于是，磁感线就成了对“电致紧张状态”作定量描绘的工具，磁感线的多少表示这种状态的强弱，磁感线的数量的变化表示这种状态的变化，这就为感生电流找到了一种量度方法，并且用“切割磁感线产生电流”对电磁感应定律做出了解释。

**5. 麦克斯韦建立电磁场理论的三篇论文**

麦克斯韦的三篇论文：1855年，《论法拉第力线》；1861年，《论物理力线》；1865年，《电磁场的动力学理论》。

第一篇文章第一部分阐述力线和不可压缩流体之间的类比。这一类比将汤姆孙的处理流线的方法作了重要的发展。把电磁学中的物理量从数学角度加以分类，提出源和旋的概念，把流体中的通量和环流迁移到电磁学。

第二部分主要讨论电磁感应现象。法拉第提出了电致紧张状态的概念，但对电磁感应没有定量表达，没有区分动生和感生电动势，麦克斯韦对法拉第提出的“电致紧张状态”进行了数学描述。1861年，他对磁场变化产生感应电动势的现象作深入分析，认识到：即使不存在导体回路，变化的磁场也会在周围激发一种场：感应电场或涡旋电场，这就区分了感生和动生。区别感生和动生有很重要的意义：把含混的感应电动力，明确改称为“涡旋电场”，发现了一种新的与静电场不同的电荷在其中受力、无源有旋的电场矢量场。场是在一定空间连续分布的，涡旋电场作为一种矢量场，不局限于某个规定的曲面周界上，把电动力改称为涡旋电场可以避免可能的误会，由变化的磁场产生的涡旋电场揭示了电场与磁场的内在联系。

第二篇文章《论物理力线》，分四部分分别载于1861年和1862年的《哲学杂志》，目的是研究介质中的应力和运动的某些状态的力学效果，并将它们与观察到的电磁现象加以比较，从而为了解力线的实质做准备。

在发表了《论法拉第力线》后，两件事情使他重新考虑研究方法，第一，法拉第力线与流体两者不宜简单类比。法拉第的力线有纵向收缩、横向扩张的趋势，力线越密，应力越大。但是流体力学中流线越密的地方压力越小，流速越快。第二，电的运动与磁的运动也无法简单类比。从电解质现象中知道电的运动是平移的，从偏振光在透明晶体中的旋转现象看，磁的运动却好像是介质中分子的旋转运动。

于是，在《论物理力线》这篇论文中，麦克斯韦提出了一个借用 W. 汤姆孙(开尔文)提出的分子涡旋模型，提出了"电磁以太模型"，如图 7-5-2 所示。图中的小球代表电以太，受电力的作用会移动，形成电流。六角形表示磁以太，绕磁感线旋转成右手关系。两者像齿轮一样互相啮合。麦克斯韦利用这个模型成功地说明了许多电磁现象。此外，磁以太作为一种弹性媒质，可以传播速度为 $v=\sqrt{\dfrac{m}{\rho}}$ 的横波($m$：切变模量，$\rho$：密度)，麦克斯韦建立了 $m$、$\rho$ 与介电常数 $\varepsilon$ 和磁导率 $\mu$ 的关系，得出真空中电磁传播速度等于光速，光波就是电磁波。

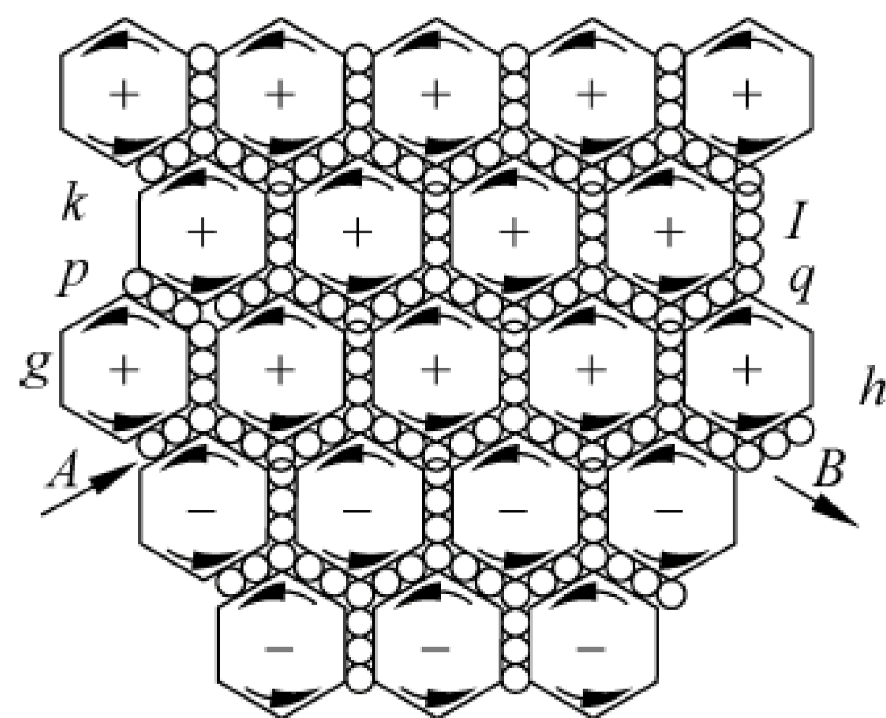


图 7-5-2 麦克斯韦"电磁以太模型"

在他的第三篇论文《电磁场的动力学理论》中，明确宣告他提出的理论可以称为"电磁场的理论"，并给出了 20 个方程 20 个变量。经过欧拉、赫兹等人的整理后，麦克斯韦方程组成了现代形式的方程积分形式和微分形式：

$$\begin{cases}\oiint \boldsymbol{D}\cdot \mathrm{d}\boldsymbol{S}=\sum q_0\\ \oiint \boldsymbol{B}\cdot \mathrm{d}\boldsymbol{S}=0\\ \oint_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=-\iint \dfrac{\partial \boldsymbol{B}}{\partial t}\cdot \mathrm{d}\boldsymbol{S}\\ \oint_L \boldsymbol{H}\cdot \mathrm{d}\boldsymbol{l}=\sum I_0+\iint \dfrac{\partial \boldsymbol{D}}{\partial t}\cdot \mathrm{d}\boldsymbol{S}\end{cases}\qquad \begin{cases}\nabla\cdot \boldsymbol{D}=\rho\\ \nabla\cdot \boldsymbol{B}=0\\ \nabla\times \boldsymbol{E}=-\dfrac{\partial \boldsymbol{B}}{\partial t}\\ \nabla\times \boldsymbol{H}=\boldsymbol{J}_0+\dfrac{\partial \boldsymbol{D}}{\partial t}\end{cases}$$

**6. 赫兹验证电磁波**

麦克斯韦理论包含的深刻和新颖的物理思想和哲学思想，一时还难以被物理学家们所理解。要人们去接受它，还需要有更强有力的证据。1870 年以后，亥姆霍兹(Helmholtz, Hermann von，1821—1894)尝试着去统一当时有的诺依曼(Neumann，Carl Gottfried von，1832—1925)、韦伯(Wilhelm Eduard Weber，1804—1891)和麦克斯韦的电磁理论。通过几年的研究他发现，关键的问题是需要证明"位移电流"的存在。1879 年，柏林科学院以"用实验建立电磁力和绝缘体介质极化的关系"为题，设立了有奖征文，要求证明以下三个假设：①如果位移电流存在，必定会产生磁效应；②变化的磁力必定会使绝缘介质产生位移电流；③在空气或真空中，上述两个假设同样成立。

赫兹是亥姆霍兹的学生。1885 年，赫兹利用一个具有初级和次级两个绕组的振荡线圈进行实验，偶然发现当初级绕组中输入一个脉动电流时，次级绕组两端的狭缝中间便会产生电火花。调整两个绕组的相对位置，火花的大小也随之变化，在某些位置下次级绕组则不产生火花。赫兹立即想到，这可能是一种电磁共振现象；既然初级绕组中的振荡电流能够激

起次级绕组的电火花,那么它就也能在邻近的介质中产生振荡的位移电流,这个位移电流又会反过来影响次级绕组的电火花发生的强弱变化。

1886 年,赫兹设计了一种直线开放振荡器:两根直导线接近的内端形成火花隙,外端则分别焊上金属板以增加电容。赫兹利用这个装置代替前述初级绕组,于同年 12 月成功地显示了直线电流引起的共振现象。

1887 年,赫兹在直线振荡器的基础上设计出了"感应平衡器"。如图 7-5-3 所示,他把一根导线弯成环状。两端留有间隙,把它放在直线振荡器 $AB$ 附近作为感应器。当将脉动电流输入 $AB$ 并在间隙处跳过火花时,在 $C$ 的间隙中也产生火花。

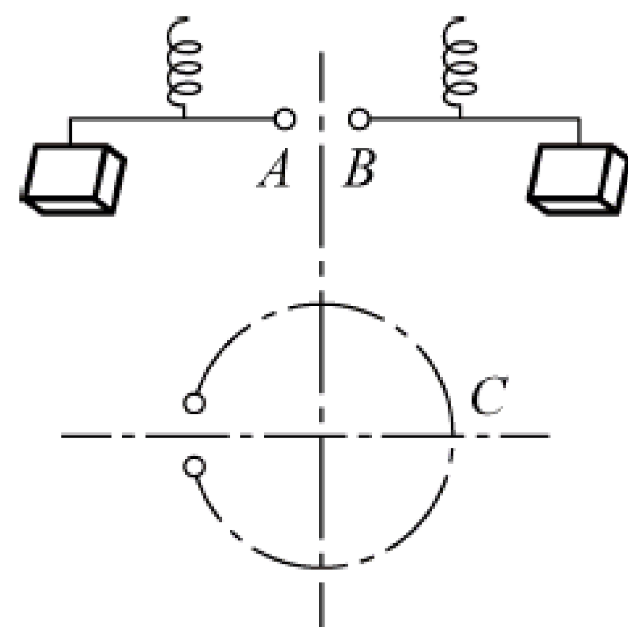



图 7-5-3 赫兹实验原理

1888 年 3 月,赫兹首先对电磁波的速度进行了测定。由于直接测定电磁波的速度是十分困难的,赫兹就利用电磁波形成的驻波测定相邻两个波节间的距离(半波长),再结合振荡器的频率计算出电磁波的速度。他在一个大屋子的一面墙上钉上一块锌皮,用来反射电磁波形成驻波。在相距十三米的地方用一个直线振荡器作为波源。赫兹用一个感应线圈作为检验器,沿驻波方向前后移动,在波节处检验器不产生火花,在波腹处产生的火花最强。赫兹利用这个方法测出了两个波节之间的长度,从而确证了电磁波的速度等于光速。

接着,赫兹又成功地用金属面使电磁波作了 45°的反射,用金属凹面镜使电磁波聚焦,用金属栅使电磁波发生偏振,以及通过用非导体材料制成的大棱镜使电磁波发生折射等,从而证实了麦克斯韦光的电磁理论的正确性。

赫兹检验电磁波以后,电磁波的实用价值引起一些工程界人士的极大兴趣。经过多年的实验,俄国科学家波波夫(1859—1906)首先获得成功。1896 年,在一次科学会议上,他和助手一起表演了传送莫尔斯电码的无线电通信实验,他们传递的电文是:"海因里兹·赫兹",通信距离是 250m。1899 年,意大利的马可尼后来居上,超过了波波夫。他成功地实现了英法海峡两岸之间的通信。1901 年,马可尼又成功地实现了跨越大西洋的无线电通信,电磁波传递的莫尔斯电码清晰地在 3600km 以外的大西洋彼岸收到。马可尼由于在无线电通信技术方面的贡献获得了 1909 年的诺贝尔物理学奖。

**7. 量子场论**

实验已经表明,以太是不存在的,于是,人们就认为物质有两种基本形态:实物和场。

我们知道,物理学的发展,使实物和场走向统一。在微观上,实物与其相应的相互作用场之间事实上不能截然分开,因为实物与场的明确界面无法确定,例如,质子和中子总为一个属于其结构的虚 $\pi$ 介子外壳所包围。因此,爱因斯坦说:"实物与场的区别,与其说是定性的问题,不如说是定量的问题。""实物便是能量密度特别大的地方,场便是能量密度小的地方"。爱因斯坦提出了一个大胆的假设,从"质能相当"出发,实物与场无定性的区别,所以,可否将实物归于场,"场是唯一的实在"。

量子场论关于物质存在的基本形态的观点有:

(1) 物质存在的两种基本形态中,场是更基本的物质存在形态,粒子不过是场处于激发态的表现。

(2) 量子场可以分为两类,它们对应的场量子不同。实物粒子场:对应的场量子为实物粒子,如电子场、中子场等。相互作用场:对应的场量子为媒介粒子,如引力场、电磁场、

强相互作用场、弱相互作用场。

(3) 真空是各种量子场均处于基态的状态。

### 7.5.2 场与路的比较

电磁学的理论包括电磁场理论和电路理论两大部分。电磁场理论主要是研究电场和磁场的性质、关系和变化规律,它用场量来描述场中各点的特性和能量分布情况。电路理论研究的是线路中的电磁场,它用与场量有密切联系并且易于测量的路量来反映线路中电磁场的变化规律。

**1. 场方程的建立与路的概念有密切联系**

电磁场理论是通过麦克斯韦提出的"位移电流"和"涡旋电场"两个著名假说,把电场、磁场的实验定律加以推广、总结而建立起来的。

电流包括电荷运动形成的传导电流和由于变化的电场形成的位移电流。显然,位移电流也是路的重要概念。它的本质说明变化的电场可以激发磁场。麦克斯韦提出位移电流的假设是在研究安培环路定律在非稳恒条件下是否适用开始的。在麦克斯韦看来,要使安培环路定理在非稳恒情况下仍然适用,关键的问题是要解决电流的连续性问题,因而提出存在位移电流的假设。在位移电流假设提出后,安培环路定律就推广为全电流定律。

涡旋电场的本质是说明变化的磁场可以激发电场。开始,麦克斯韦提出涡旋电场假设是为了说明穿过闭合回路的磁通量变化时产生感应电动势的原因。在麦克斯韦看来,穿过闭合回路的磁通量发生变化时之所以在回路中产生电动势,是由于回路中存在涡旋电场(感应电场)。于是,法拉第的电磁感应定律的形式 $\varepsilon=-\dfrac{\mathrm{d}\Phi}{\mathrm{d}t}$就被推广为

$$\varepsilon=\frac{\mathrm{d}\Phi}{\mathrm{d}t}=-\iint\frac{\partial \boldsymbol{B}}{\partial t}\cdot\mathrm{d}\boldsymbol{S} \tag{7-5-1}$$

这是涡旋电场的环流定理,它与库仑场的环路定理合在一起就变成为

$$\oint\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{l}=-\iint\frac{\partial \boldsymbol{B}}{\partial t}\cdot\mathrm{d}\boldsymbol{S} \tag{7-5-2}$$

这是场的重要方程之一,它的建立与电动势有着密切的联系。

**2. 基本场量与基本路量之间的关系**

描述电磁场特征的基本物理量为:电场强度、磁场强度和电位移矢量,分别用 $\boldsymbol{E}$、$\boldsymbol{H}$ 和 $\boldsymbol{D}$ 表示;而描述电路特征的基本物理量为:电流、电压和电荷,通常用 $i$、$u$ 和 $q$ 表示。电磁场理论是研究在无限伸展的三维空间中各点所发生的电磁现象,而电路理论则是研究在一个特定的局部空间所发生的电磁现象。从理论研究的意义上来说场论和路论是研究电磁现象的两种不同的科学观点和科学方法。因此,场的表征量与路的表征量之间必然存在着相互联系,它们之间的对应关系如下:

$$\begin{cases} i=\oint\boldsymbol{H}\cdot\mathrm{d}\boldsymbol{l} \\ u=\oint\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{l} \\ q=\oiint\boldsymbol{D}\cdot\mathrm{d}\boldsymbol{S} \end{cases} \tag{7-5-3}$$

由此可以看出:场量与路量之间不是彼此孤立的,而是相互联系的。如电路中的电压是指沿某一途径电场强度的线积分 $u=\oint\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{l}$,而电路中的电流,则表示围绕电路导线截面磁场

强度的线积分 $i=\oint \boldsymbol{H}\cdot \mathrm{d}\boldsymbol{l}$，同时它们也有各自的特点，场量描绘的是场中各点的电磁特性，它们不仅是时间的函数，也是空间的函数，是微分量、矢量；而路量描绘的是总的特性，它们仅为时间函数(分布参数电路例外)，是积分量、标量。

**3. 稳恒电路的欧姆定律**

在稳恒电路中，若外电阻是一段导体，当两端加上电压 $U$ 时，电流强度与电压成正比，即 $I=\dfrac{U}{R}$，$R$ 叫电阻，上式是稳恒电路的欧姆定律。对于均匀导体，若其长度为 $L$，$S$ 为导体的横截面积，$\sigma$ 为导体的电导率，则可以得到 $I=\dfrac{U}{R}=\sigma\dfrac{U}{L}S$。由于 $jS=\sigma\dfrac{U}{L}S$，即

$$\boldsymbol{j}=\sigma\boldsymbol{E} \tag{7-5-4}$$

如果推广到电源区，则 $\boldsymbol{j}$ 由 $\boldsymbol{E}$ 和外场 $\boldsymbol{k}$ 共同决定，有 $\boldsymbol{j}=\sigma(\boldsymbol{E}+\boldsymbol{k})$，这就是全电路的欧姆定律的微分形式。

**4. 基尔霍夫定律**

欧姆定律只能用于解比较简单的电路，复杂的电路，有许多条导线交汇于一点，整个电路由若干个闭合回路组成，同一回路的各段电路中的电流并不相同。对于这类复杂电路，欧姆定律无法解决，这时候，就得用到基尔霍夫定律。

在复杂电路中，往往有几条支路交会于一点。凡是三条或三条以上支路的交汇点，称为节点，根据电荷守恒定律和稳恒电流条件，流入分支点的电流应等于流出分支点的电流，因此，对于每一个节点，有

$$\sum I_k=0 \tag{7-5-5}$$

$I_k$ 是通过节点的各支路的电流。这就是基尔霍夫(Kirchhoff Gustav Robert，1824—1887)第一定律，也称为节点(电流)定律。

基尔霍夫第一定律可以从稳恒电场的特性去推出。由 $\oiint \boldsymbol{j}\cdot \mathrm{d}\boldsymbol{S}=0$ 出发，对于任一节点来说，$\oiint \boldsymbol{j}\cdot \mathrm{d}\boldsymbol{S}=\sum I_k=0$，即汇集于任何一个节点的电流强度代数和为零。对于复杂电路中任一回路，我们可以运用非均匀电路欧姆定律计算回路上任意两点电势差，若计算电势 $a$、$b$ 两点重合，则有 $\varphi_a-\varphi_b=\sum IR-\sum\varepsilon=0$，即得

$$\sum IR=\sum\varepsilon \tag{7-5-6}$$

即沿任意闭合回路绕行一周，回路中各电阻上电势降的代数和等于各电源的电动势造成电势升高的代数和，这就是基尔霍夫第二定律，也称为基尔霍夫电压定律。它的基础是稳恒电场的环路定理：$\int_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=0$。所以，基尔霍夫第二定律也可以从稳恒电场的特性去推出。

根据环路定理，沿回路环绕一周回到出发点，电势没有改变：$0=\int_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=\sum IR-\sum\varepsilon$，即得 $\sum IR=\sum\varepsilon$。

从上面的讨论可以看出，电路的基尔霍夫定律实质上是电磁场基本规律的简化，说明了路论和场论是统一的。

**5. 由路方程推出麦克斯韦方程**

麦克斯韦方程可以认为是路方程的推广。

从路的角度，由安培环路定律 $\int_L \boldsymbol{H} \cdot \mathrm{d}\boldsymbol{l} = \sum I$ 出发，如果再考虑到位移电流，可以得到 $\int_L \boldsymbol{H} \cdot \mathrm{d}\boldsymbol{l} = \int_L \left(\boldsymbol{j} + \frac{\partial \boldsymbol{D}}{\partial t}\right) \cdot \mathrm{d}\boldsymbol{S}$，该式的微分形式即是 $\nabla \times \boldsymbol{H} = \boldsymbol{j} + \frac{\partial \boldsymbol{D}}{\partial t}$，这就是场的方程。

从路的角度，电动势的定义式 $\varepsilon = \int_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l}$，由法拉第电磁感应定律：$\varepsilon = -\frac{\mathrm{d}\Phi}{\mathrm{d}t}$，应用 $\boldsymbol{E}$ 线积分和 $\boldsymbol{B}$ 面积分表示，则有 $\int_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\iint_S \frac{\partial \boldsymbol{B}}{\partial t} \cdot \mathrm{d}\boldsymbol{S}$，这就是场的方程。

此外，还可根据高斯定理导出电场和磁场方程：$\nabla \cdot \boldsymbol{D} = \rho$ 和 $\nabla \cdot \boldsymbol{B} = 0$。

由路方程推出麦克斯韦方程，再次证明了路论与场论是统一的。

**6. 场论的普遍性与路论的局限性**

电磁场理论，是通过研究电场强度、磁感应强度等场量在空间的分布来研究物理系统内发生的电磁过程的。因此，场的方法更严谨、更具普遍性。原则上几乎所有的电磁问题都可以用麦克斯韦方程及其边界条件来解决。然而，实际上只有在边界情况比较简单的情况下，才能求得严格的解，而且求解过程也非常繁琐。

当电源电动势或电压作周期性变化时，尤其是在迅变情况下，例如，电台天线上的高频振荡电流，天线激发的电磁场能脱离天线以波的形式向远方传播，在真空中的传播速度为 $c = 3 \times 10^8 \mathrm{m/s}$，由 $\lambda = c/f$ 可知，频率越高，波长就越短，当电路的尺寸与波长相等或者波长小于电路的尺寸时，在电路上的不同地方，电场强度不同，因此，基尔霍夫电流定律不适用了。同时，在这种迅变的电磁场和电路中，电场强度的环流显然不为零，不能像稳恒电场那样引入电势差或电压的概念，所以，基尔霍夫电压定律自然也就不适用了。可见，路的方程是有局限性的。

电路研究方法虽有其局限性(它是场的问题在一定条件下的近似)。然而在求解某些具体问题时能取到简化的作用。例如，对一个简单的 $R$、$L$、$C$ 串联电路，要求出一个满足所有边界条件的电磁场的解，就不是很容易的事，但采用电路分析的方法来分析该电路，便要简单得多。对于一个具体的问题，我们应根据实际情况，做出具体的分析，以便对这两种方法做出合适的选择。一般来说，当某物理系统各方向的尺寸远小于所传递的电磁波的波长时(即通常所说的似稳条件)，则可按集中参数电路的观点进行分析，因为在这种情况下，可以忽略电磁场的推迟效应，对同一条无分支的导线来讲，可近似地认为在同一时刻电流是相同的，同时由于在似稳条件下，信号源的频率较低，电路中的涡旋电场可以忽略不计，才能引入电压的概念。而对某些工程实际问题，如电工设备中的绝缘耐压、大地导体中电流及磁通的分布、电磁干扰及屏蔽、电磁波的辐射及传播等问题同时发生必须用电磁场的方法来分析与计算。

## 7.6　麦克斯韦方程组

麦克斯韦方程组诞生的关键是“位移电流”的思想实验，本节从麦克斯韦方程组的产生、来源历史、表述形式等方面作了比较详细的讨论，加深对麦克斯韦方程组的认识以及对电磁学理论有完整的理解。

麦克斯韦方程组的建立可以说是19世纪的物理学上登峰造极的成就。著名物理学家费曼(Richard Phillips Feynman，1918—1988)说：“从人类历史的漫长远景来看，毫无疑问，

在19世纪中发生的最有意义的事件将判定是麦克斯韦对电磁定律的发现。”麦克斯韦方程组指明了电磁场运动变化所遵从的基本规律,它和洛伦兹力公式以及电荷守恒定律一起构成了经典电磁现象的完整理论基础。尽管在高速运动的条件下要考虑电磁场的变换关系,在微观领域里要考虑量子化效应,但作为电磁场的普遍规律的麦克斯韦方程组,其形式仍然成立。因此,麦克斯韦方程组的诞生是物理学具有划时代的一次大统一。

### 7.6.1 麦克斯韦方程组的实验基础

电磁学定律是从电学实验中发现和总结出来的,当时发现(恒定的)电流可以产生(恒定的)磁场,这主要由安培环路定律表示,但是恒定的磁场却不会产生电流;另一方面,变化的磁场才可以导致电流的产生,这主要由法拉第电磁感应定律表示,人们却没有与之对应的变化的电场的概念,这种电磁关系的不对称并没有引起当时实验物理学家的特别关注,因为在当时的实验条件下看不到这些现象。

但是,麦克斯韦的工作不同,他完全用数学语言来表达电磁定律,从而使这种不对称的缺陷充分暴露出来,但是麦克斯韦并不能直接从这种不对称性关系中推演出对称性来,他仍然只能回到实验中去,不同的是他不用做实验室中的实验,他只需做思想实验,这种思想实验不是数学表达式在思想中的推演,而是在思想中进行的对电和磁的运动形象过程的再创造。他沿用安培定律的实验,想象电流和磁场的运动过程,当时的情形在现在看来是非常奇特的,传统的物理学家只能沿用经典图像进行思考(甚至今天在大多数情况中也只能这样)。比如,把电和磁想像为以太流体、涡旋、弹性物质,甚至齿轮之类,麦克斯韦的思想实验也是在这样的图像中进行的。但是,由于麦克斯韦脱离了具体实验环境的限制,所以,他能在他的思想实验中“观察”到新的“现象”。

麦克斯韦工作的关键是他的著名的所谓“位移电流”的思想图像,即把变化的电场也看成为一种(以太)电流。事实上,电场在物理过程上可以解释为电介质内的分子产生极化的状态,它是分子中的外层电的总的位移效应(当时还没有发现“电子”!),在当时的实验室条件下观察不到这种效应所表现出来的现象,而位移电流是一个在思想实验中的能够被“观察”到的交变电流过程:你可以想象有一种流态的电物质在物质中来回移动(交变电流)而不是单向通过(稳恒电流),这样它就脱离了实验室条件下具体的导体或绝缘体的物理限制,使电场能以电流的形象出现,这种交变的位移电流产生交变磁场,这样交变的电、磁场可以相互产生,电与磁的对称性成为在理论表达上完全具有共性的一种本质关系,这种在相互转化的对称性中的电与磁的统一就是电磁场。位移电流的思想实验,直接导致麦克斯韦在以前的安培环路公式中添加电场的变化率一项,这就是麦克斯韦方程组物理本质化的一个关键。这样,麦克斯韦就成功地把静态意义的安培环路公式改造成了交变(电磁场)的安培环路公式,奠定了电磁场数学表达形式在本质上的统一,使以前没有内在共同统一性的静电学和静磁学转变成为电磁场理论的电动力学。

由此,我们可以看出,麦克斯韦并不是完全靠数学演绎方法直接从库仑定律、安培环路定律、法拉第电磁感应定律等数学表达式中推导得到了麦克斯韦方程组,也不是由既定演绎性前提推导出新的结果,而是首先用思想实验方法发现了安培环路定律的新的意义,补充了安培环路公式,从而揭示了电与磁的物理现象后面存在的共同的本质,这样才使以前几个相互没有内在统一性的电磁公式成为具有本质性意义相互联系的麦克斯韦方程,成为可以表达一种全新的物理对象的数学形式。

## 7.6.2　麦克斯韦方程组的思维方法

从麦克斯韦方程组的产生知道，它的产生和推广的关键有两个方面：其一是假定变化的电场(位移电流)产生磁场，从而把安培环路定理加以推广，使之包括位移电流；其二是假定变化的磁场产生电场，从而把法拉第电磁感应定律由导体回路中产生感应电动势推广到一般情况，提出了涡旋电场的概念。麦克斯韦方程的来源如图 7-6-1 所示。

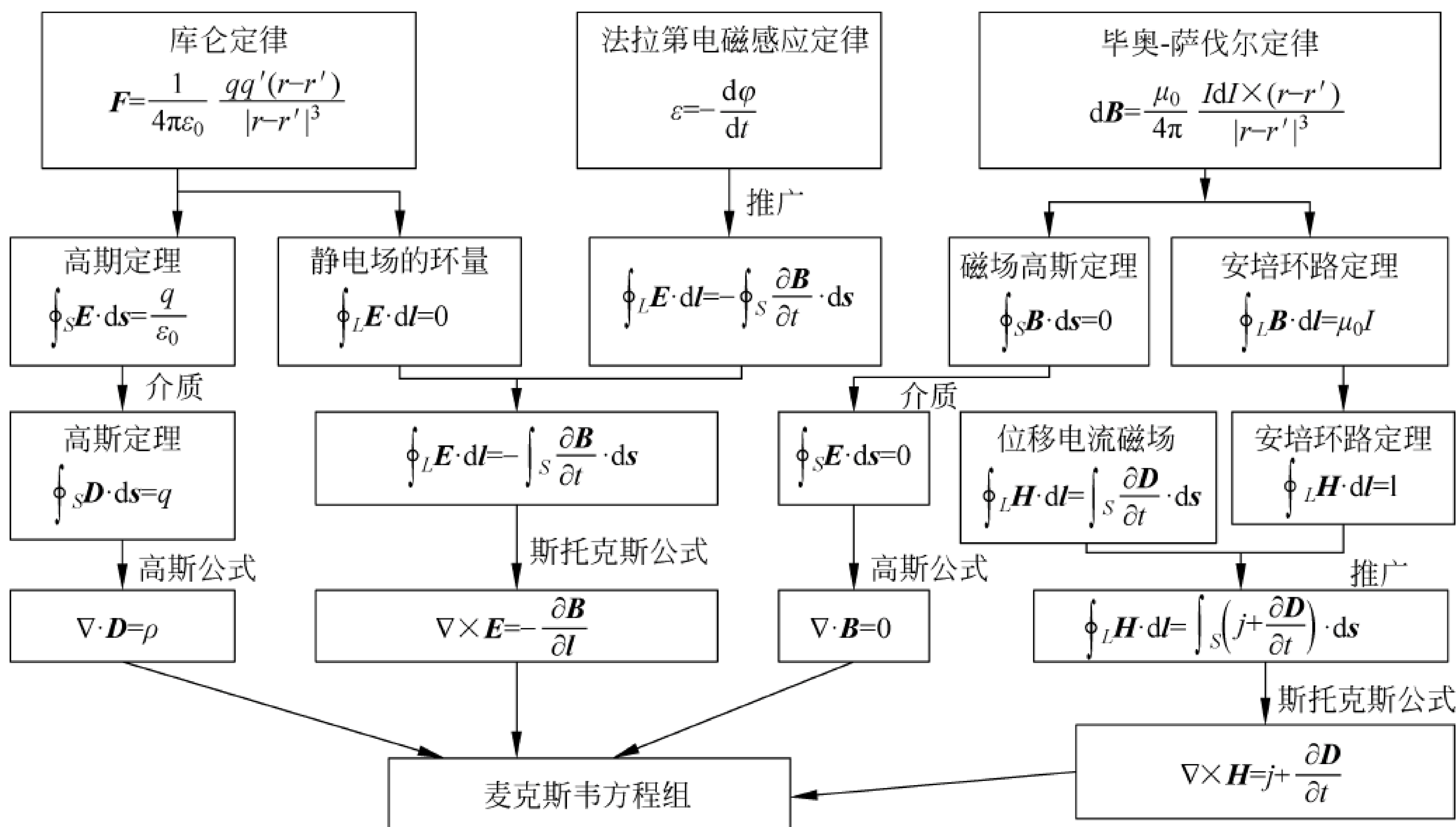



图 7-6-1　麦克斯韦方程的来源示意图

## 7.6.3　麦克斯韦方程组的建立过程

### 1. 稳恒场中麦克斯韦方程的建立

稳恒电场的高斯定律为 $\int \boldsymbol{D} \cdot \mathrm{d}\boldsymbol{s} = q_0$，如果 $q_0$ 呈连续分布，就有

$$\oint_S \boldsymbol{D} \cdot \mathrm{d}\boldsymbol{s} = \oint_V \rho_0 \, \mathrm{d}v \tag{7-6-1}$$

运用类比思维，在稳恒磁场中，因为磁感线是闭合的，在磁场中进入某一闭合面的磁感线通量等于从这个面穿出的通量，于是得到磁场的高斯定律：

$$\oint \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{s} = 0 \tag{7-6-2}$$

安培环路定律为 $\oint \boldsymbol{H} \cdot \mathrm{d}\boldsymbol{l} = I_0$，若 $I_0$ 以电流密度 $J_0$ 在某一截面上连续分布，则有

$$\oint_L \boldsymbol{H} \cdot \mathrm{d}\boldsymbol{l} = \oint_S \boldsymbol{J}_0 \cdot \mathrm{d}\boldsymbol{s} \tag{7-6-3}$$

如果磁通量的变化是由于磁场随时间 $t$ 变化引起的，即 $\boldsymbol{B}=\boldsymbol{B}(\boldsymbol{r},t)$，那么，变化的 $\boldsymbol{B}$ 在某一回路中就激发出感应电动势 $\varepsilon$，其定量关系由法拉第电磁感应定律给出，$\varepsilon = -\dfrac{\mathrm{d}\varphi}{\mathrm{d}t}$，而 $\varphi = \int \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{s}$，则有 $\varepsilon = \dfrac{\mathrm{d}\varphi}{\mathrm{d}t} = -\int \dfrac{\partial \boldsymbol{B}}{\partial t} \cdot \mathrm{d}\boldsymbol{s}$，而 $\varepsilon = \oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l}$，即电动势产生意味着一定存在某种

非静电力在做功，也就是一定存在某种非静电场，麦克斯韦假定在变化的磁场周围激发出这种电场 $\boldsymbol{E}$，它和静电场共同的是对电荷有作用力，不同的是这种场不是电荷激发的，而且，描述这种场的力线是闭合的，这种场称为感应电场或涡旋电场。它存在于导体、介质、空间中，在闭合导体中形成电流。因此，感应电动势 $\varepsilon$ 是感应电场强度 $E$ 沿闭合路径的线积分。

$$\oint \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=-\int \frac{\partial \boldsymbol{B}}{\partial t}\cdot \mathrm{d}\boldsymbol{s} \tag{7-6-4}$$

式(7-6-1)～式(7-6-4)是麦克斯韦方程组在稳恒场中的积分形式，为了逐点研究场，就要将积分形式变为微分形式。根据高斯散度定理 $\oint_S \boldsymbol{A}\cdot \mathrm{d}\boldsymbol{s}=\int_V \nabla\cdot \boldsymbol{A}\,\mathrm{d}v$，方程(7-6-1)写为 $\oint_S \boldsymbol{D}\cdot \mathrm{d}\boldsymbol{s}=\int_V \nabla\cdot \boldsymbol{D}\,\mathrm{d}v=\int_V \rho_0\,\mathrm{d}v$，则有

$$\nabla\cdot \boldsymbol{D}=\rho_0 \tag{7-6-1$'$}$$

同理，方程(7-6-2)写为 $\oint_S \boldsymbol{B}\cdot \mathrm{d}\boldsymbol{s}=\int_V \nabla\cdot \boldsymbol{B}\,\mathrm{d}v=0$，则有

$$\nabla\cdot \boldsymbol{B}=0 \tag{7-6-2$'$}$$

根据斯托克斯定理 $\oint_L \boldsymbol{A}\cdot \mathrm{d}\boldsymbol{l}=\oint_S \nabla\times \boldsymbol{A}\cdot \mathrm{d}\boldsymbol{s}$，方程(7-6-3)写为 $\oint_L \boldsymbol{H}\cdot \mathrm{d}\boldsymbol{l}=\int_S \nabla\times \boldsymbol{H}\cdot \mathrm{d}\boldsymbol{s}=\int_S \boldsymbol{J}_0\cdot \mathrm{d}\boldsymbol{s}$，则有 $\nabla\times\boldsymbol{H}=\boldsymbol{J}_0$。

方程(7-6-4)写为 $\oint_L \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}=\int_S \nabla\times \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{s}=-\int_S \frac{\partial \boldsymbol{B}}{\partial t}\cdot \mathrm{d}\boldsymbol{s}$，则有

$$\nabla\times \boldsymbol{E}=-\frac{\partial \boldsymbol{B}}{\partial t} \tag{7-6-4$'$}$$

以上四式是麦克斯韦方程组在稳恒场中的微分形式。

**2. 时变场中麦克斯韦方程的建立**

麦克斯韦是怎样把稳恒场方程推广到时变场的呢？基本原则是：理论上应当自洽，实验上不冲突。$\nabla\cdot\boldsymbol{D}=\rho_0$ 和 $\nabla\cdot\boldsymbol{B}=0$ 是适用于静电场和稳恒电流的磁场，然而，由 $D(v)$、$B(v)$ 过渡到 $D(v,t)$、$B(v,t)$，在理论上不矛盾，也未发现它们与以后的实验相冲突，因此，就将 $\nabla\cdot\boldsymbol{D}=\rho_0$ 和 $\nabla\cdot\boldsymbol{B}=0$ 两方程直接推广到时变场中。法拉第电磁感应定律 $\nabla\times\boldsymbol{E}=-\frac{\partial \boldsymbol{B}}{\partial t}$ 本身就是在时变场中建立的，麦克斯韦认为环路定律 $\nabla\times\boldsymbol{H}=\boldsymbol{J}_0$ 对时变场不成立，原因是在时变场中 $\nabla\times\boldsymbol{H}=\boldsymbol{J}_0$ 与电荷守恒定律相矛盾。电荷守恒定律 $\int_S \boldsymbol{J}_0\cdot \mathrm{d}\boldsymbol{s}=\int_V \nabla\cdot \boldsymbol{J}_0\,\mathrm{d}v=-\frac{\mathrm{d}q}{\mathrm{d}t}=-\int_V \frac{\mathrm{d}\rho_0}{\mathrm{d}t}\mathrm{d}v$，于是就有

$$\nabla\cdot \boldsymbol{J}_0=-\frac{\mathrm{d}\rho_0}{\mathrm{d}t} \tag{7-6-5}$$

对 $\nabla\times\boldsymbol{H}=\boldsymbol{J}_0$ 两端取散度 $\nabla\cdot(\nabla\times\boldsymbol{H})=\nabla\cdot\boldsymbol{J}_0$，由场论知识 $\nabla\cdot(\nabla\times\boldsymbol{H})=0$，则有

$$\nabla\cdot \boldsymbol{J}_0=0 \tag{7-6-6}$$

显然式(7-6-5)与式(7-6-6)相矛盾，因为电荷守恒定律是普遍成立的，所以式(7-6-5)是正确的。

麦克斯韦是怎样修改 $\oint_L \boldsymbol{H}\cdot \mathrm{d}\boldsymbol{l}=I_0$ 的呢？

在如图 7-6-2 所示接入电容器的电路中，取包围正极板的闭合回路 $L$ 为周界的任意曲面 $S$ 的传导电流为 $I_0$，由于电容器处于充放电过程，$I_0$ 总随时间 $t$ 变化。以 $L$ 为周界作两个曲面 $S_1$ 与导线相交，$S_2$ 在电容器两极板间，计算通过 $S_1$ 面的电位移通量。取 $S_1$ 和 $S_2$ 在极板表面两侧，$S_1$ 在导体内，$S_2$ 在导体外，如图 7-6-3 所示。

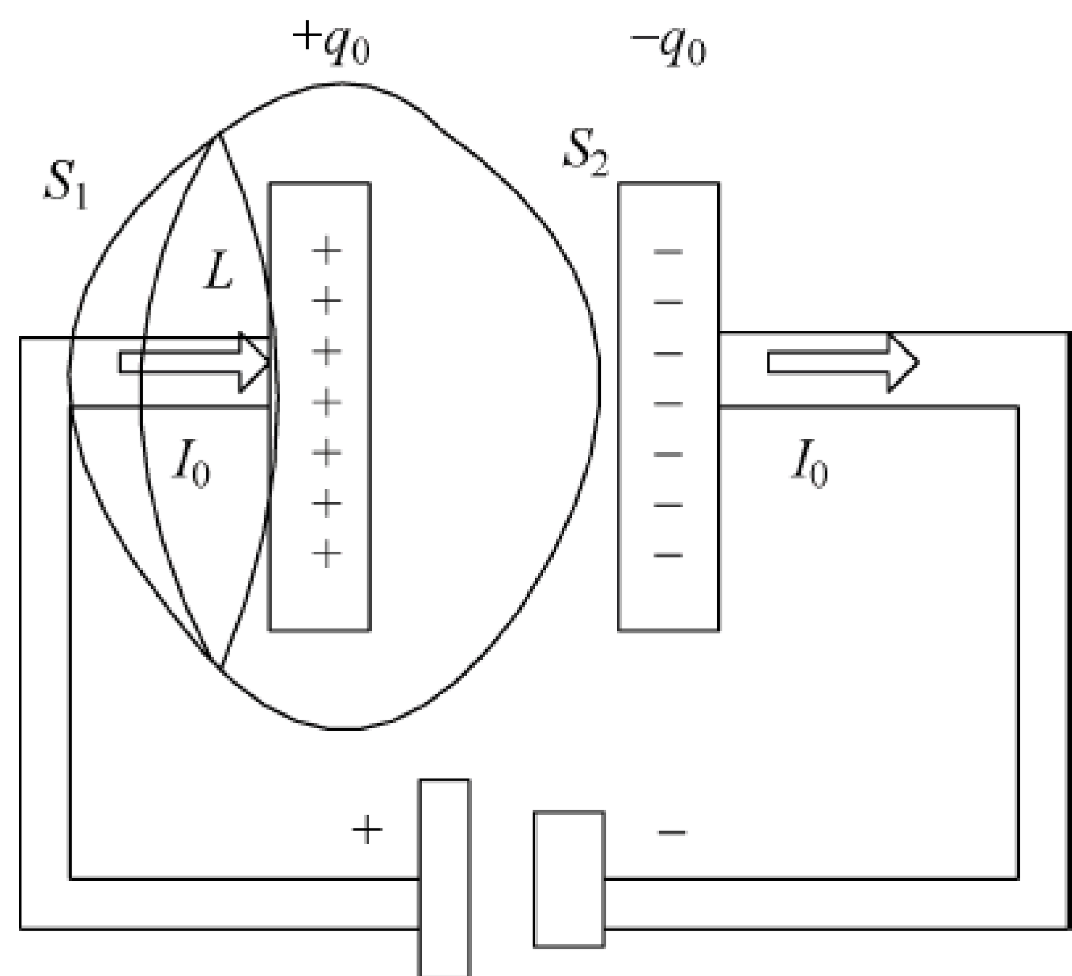



图 7-6-2 非稳恒电流的情形

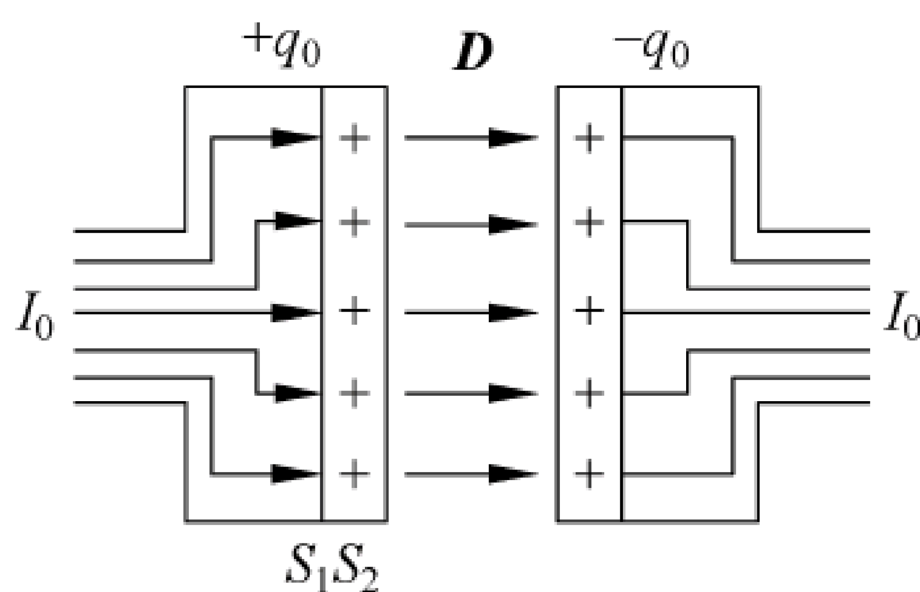



图 7-6-3 位移电流——变化的电场

根据高斯定律 $\phi_D=\oint_S \boldsymbol{D}\cdot \mathrm{d}\boldsymbol{s}=\oint_{S_1+S_2}\boldsymbol{D}\cdot \mathrm{d}\boldsymbol{s}=\int_{S_1}\boldsymbol{D}\cdot \mathrm{d}\boldsymbol{s}+\int_{S_2}\boldsymbol{D}\cdot \mathrm{d}\boldsymbol{s}=q_0$，因 $S_1$ 在导体内，故 $\boldsymbol{D}=0$，$\boldsymbol{E}=0$，于是就有 $\phi_D=\int_{S_2}\boldsymbol{D}\cdot \mathrm{d}\boldsymbol{s}=q_0$，又因 $I_0=\frac{\mathrm{d}q_0}{\mathrm{d}t}=\frac{\mathrm{d}\phi_D}{\mathrm{d}t}=\int_{S_2}\frac{\partial \boldsymbol{D}}{\partial t}\cdot \mathrm{d}\boldsymbol{s}$，因此得到 $I_0=\int\frac{\partial \boldsymbol{D}}{\partial t}\cdot \mathrm{d}\boldsymbol{s}$。说明通过 $S_1$ 的传导电流 $I_0$ 等于穿过 $S_2$ 的电位移通量的时间变化率。麦克斯韦把 $\frac{\mathrm{d}\phi_D}{\mathrm{d}t}=\int\frac{\partial \boldsymbol{D}}{\partial t}\cdot \mathrm{d}\boldsymbol{s}$ 称为位移电流。位移电流是电流的扩充，它是变化的电场，与传导电流有同样的磁效应。引入位移电流之后，在电容器表面中断的传导电流被位移电流接替下去，二者合在一起保持着电流的连续性。麦克斯韦又将传导电流 $I_0$ 和位移电流 $\frac{\partial \boldsymbol{D}}{\partial t}\cdot \mathrm{d}\boldsymbol{s}$ 之和称为全电流。稳恒场中 $\oint \boldsymbol{H}\cdot \mathrm{d}\boldsymbol{l}=I_0$ 对时变场不适用的原因是传导电流的不连续性，麦克斯韦用全电流代替传导电流保证了非稳恒情况下电流的连续性，从而把稳恒场中的环路定律推广到时变场，即普遍情况下的安培环路定律：

$$\oint \boldsymbol{H}\cdot \mathrm{d}\boldsymbol{l}=I_0+\int\frac{\partial \boldsymbol{D}}{\partial t}\cdot \mathrm{d}\boldsymbol{s} \tag{7-6-7}$$

并由此得到它的微分形式：

$$\nabla\times \boldsymbol{H}=\boldsymbol{J}_0+\frac{\partial \boldsymbol{D}}{\partial t}=\boldsymbol{J}\quad(\boldsymbol{J}\text{ 是全电流密度}) \tag{7-6-7′}$$

至此，麦克斯韦完成了电磁场方程的建立，由式(7-6-1)和式(7-6-2)、式(7-6-4)和式(7-6-7)四式组成。式(7-6-1′)和式(7-6-2′)、式(7-6-4′)和式(7-6-7′)四式是相应的微分形式。

### 7.6.4 麦克斯韦方程组的伟大意义

麦克斯韦方程组是麦克斯韦在电磁现象已有的定律及实验基础上进行分析、总结，并引入位移电流、全电流和感应电场(涡旋电场)的概念，在数学上对电磁规律作了高度的概括，于 1873 年建立起来的。全电流密度是麦克斯韦电磁场理论的核心，引入位移电流和感应电

场(涡旋电场)的概念是麦克斯韦对电磁场理论的重要发展。位移电流说明时变电场可以产生磁场,感应电场(涡旋电场)说明时变磁场可以产生电场。因此,麦克斯韦根据他建立的方程从理论上得出电磁波存在的预言和光是电磁波的论断,1887 年由赫兹用实验证明了电磁波的存在。从此,物理学开始了一个新的时期。电磁波成功地在无线电通信中的应用,充分证明了电磁场方程是正确的。光是电磁波这种论断的提出是人类认识光本性方面的一大飞跃。正是在这个意义上,人们说麦克斯韦把电学、磁学和光学统一起来了。这一发展是 19 世纪科学史上最伟大的成就,是继牛顿力学之后又一伟大的贡献。它在经典电磁学中的地位与牛顿定律在经典力学中的地位相同。麦克斯韦的电磁场理论和牛顿力学体系构成了经典物理学的两大支柱。纵观麦克斯韦建立电磁场方程的全部过程,不仅使我们得到了完美的电磁场理论,而且在科学研究方法论方面也给我们留下了宝贵的精神财富。麦克斯韦方程建立至今已百余年了,今天人们仍然折服于它的完整、系统、简洁和严密,它是人类认识史上不可多得的完美理论之一。

# 学习资源

## 一、中国近现代著名科学家(物理学)简介

邓稼先(1924—1986),中国安徽怀宁人,著名核物理学家,中国科学院院士。我国核武器理论研究工作的奠基者之一。从原子弹、氢弹原理的突破和试验成功及其武器化,到新的核武器的重大原理突破和研制试验,均做出了重大贡献。中国核武器研制与发展的主要组织者、领导者。邓稼先历任中国科学院近代物理研究所助研、副研究员,二机部第九研究所理论部主任、第九研究院副院长、院长,国防科工委科技委副主任,核工业部科技委副主任等职。邓稼先还在中国共产党第十二次全国代表大会上被选为中央委员。邓稼先始终在中国武器制造的第一线,领导了许多学者和技术人员,成功地设计了中国原子弹和氢弹,把中华民族国防自卫武器引领到了世界先进水平。邓稼先于 1967 年领导亲自参与了中国第一颗氢弹的研制和实验工作。1982 年获国家自然科学奖一等奖,1985 年获两项国家科技进步奖特等奖,1986 年获全国劳动模范称号,1987 年和 1989 年各获一项国家科技进步奖特等奖。1999 年,中共中央、国务院、中央军委追授他“两弹一星”功勋奖章。邓稼先同志在一次实验中,受到核辐射,身患直肠癌,于 1986 年 7 月 29 日在北京不幸逝世,终年 62 岁。



## 二、诺贝尔物理学奖简介

**1951 年诺贝尔物理学奖**。通过人工加速的粒子轰击原子,促使其产生核反应(嬗变)。获奖者:J. D. 科克罗夫特(英国)、E. T. S. 沃尔顿(爱尔兰)。评价:此二人最大的贡献就是改善了加速器使粒子能获得更快的速度,并用这种新型的加速器实现了人类第一次完全由人工控制的方法使原子核发生嬗变。嬗变即原子核变成另一种原子核。所谓工欲善其事,必先利其器。从前面的很多例子可以看出,重大的研究成果离不开对实验器材的改造,而对器材的成功改造正是一个国家综合国力的体现。

**1952 年诺贝尔物理学奖**。发展核磁精密测量的新方法及其有关的发现。获奖者:F. 布洛赫、E. M. 珀塞尔(美国)。评价:他二位的主要贡献就是分别用不同的方法独自提高

了测量原子核磁矩的精度，对核磁共振技术的发展起到了重要作用。对人类认识微观世界，治愈疾病有重要作用。此外布洛赫提出了能带的理论，从物质的结构上阐明了导体、半导体、绝缘体的区别。

**1953 年诺贝尔物理学奖**。发明了相衬显微镜。获奖者：F. 泽尔尼克（荷兰）。评价：由于人类的眼睛不能辨别物体的相位变化（即根据光的波动理论 $E=A\sin(ax+b)$），人只能辨别 $a$ 不能观察 $b$，所以以前的显微镜只能观察不透明的物体。对于一般显微镜来说只有染色才可以观察。所以人们无法直接观察活细胞一直是生物学的一大遗憾。但泽尔尼克却想办法将光的相位变化的信息转化为振幅变化的信息。这样人眼就可以识别了。实现人类观察未经染色物体和活细胞的梦想。

**1954 年诺贝尔物理学奖**。在量子力学和波函数的统计解释及研究方面做出贡献，发明了符合计数法，用以研究原子核反应和 γ 射线。获奖者：W. 博特（德国）、M. 玻恩（德国）。评价：自从德布罗意提出物质波以来，人们就在争论物质波的实质是什么。比如有人设想波是大量粒子的密度变化产生的一种密度波。又有人如薛定谔认为是一种波包（这个还是上维基自己看吧），但是理论上又出现了明显的漏洞。而玻恩提出了所谓粒子的波动性实际上是它概率分布的波动性。即在某些地方出现的可能性大另一些地方可能性小。而他的概率波理论和海森伯的不确定性关系，玻尔的互补性原理构成了量子力学的正统解释——哥本哈根解释。虽然哥本哈根解释在各方面应用取得了巨大成功，但由于过于重视数学推理，忽略了是否能在哲学上说得通，所以自提出起就遭到了很多批评与探讨。很多反对者都是对量子力学功勋卓著的人，如爱因斯坦，薛定谔，德布罗意。这些批评一方面让人们能对量子力学有更深的认识，比如现在非常热门的量子信息理论就是这中间提出来的。博特的主要贡献是提出了一种新的实验方法——符合计数法，并由此证明能量和动量守恒定律在光子和电子碰撞中依然有效。1930 年前后，宇宙线领域里的一些重要发现几乎都和符合计数法分不开。

**1955 年诺贝尔物理学奖**。发明了微波技术，进而研究氢原子的精细结构，用射频束技术精确地测定出电子磁矩，创新了核理论。获奖者：P. 库什（美国）、W. E. 拉姆（美国）。评价：他们两位主要是对微观世界结构有了进一步的发展。氢原子是最简单的原子，而人类通过对它的研究掌握了很多关于原子的基础知识。他们通过实验让人们对此有了新的认识，测量出了在量子力学中几个十分有用的常数。

**1956 年诺贝尔物理学奖**。从事半导体研究并发现了晶体管效应。获奖者：W. H. 布拉顿、J. 巴丁、W. B. 肖克利（美国）。评价：晶体管的发明可以说是 20 世纪物理学的一件大事，标志着现代电子学的诞生。大家使用电脑时不要忘了他们三位的功绩。晶体管是现代电路的基本元件，是构成数字电路的基础。所有的电子产品无不以晶体管为基础（虽说现在使用 mos 管技术三极管用的很少，虽然内部构造不同但它们的原理还是一样的）。而现代社会也因此以摩尔定律的速度成几何级数向前发展。可以说 20 世纪后半叶信息时代全球化。数字化带来的生产关系的社会巨变离不开他们三位对生产力的发展。

**1957 年诺贝尔物理学奖**。发现在弱作用过程中宇称不守恒。获奖者：李政道、杨振宁（中国）。评价：这次的获奖是值得我们中国人骄傲的一件事，证明了人类的最高智力殿堂也有我们的一席之地。要解释他们的理论首先说明物理定律有三个守恒：其一，电荷守恒；其二，宇称守恒；其三，电子守恒。其中宇称守恒是说对于一个只有自旋方向相反的粒子其他什么都相同（形象的比喻就是照镜子），那么物理定律对于二者是相同的。但当时实验却

出现了问题：实验上发现的 $q$ 粒子与 $t$ 粒子具有几乎完全一样的性质——相同的质量、相同的寿命、相同的自旋、相同的电荷……以至于人们不得不怀疑它们是否就是同一个粒子，然而它们在衰变时却表现出不同的性质。后来杨振宁、李政道两人证明，宇称守恒定律在弱相互作用下不成立。即那两个粒子由于自旋方向相反，所以在弱相互作用下应表现出不同的性质。这项发明对粒子物理学是场革命。1957 年，拉比(1944 年诺奖得主)评论道："就某种意义上说，一个相当完善的理论结构从根本上被摧垮了，我们不能肯定这些碎片将怎样再组合在一起。"

**1958 年诺贝尔物理学奖**。发现并解释了切伦科夫效应。获奖者：P. A. 切伦科夫、I. E. 塔姆、I. M. 弗兰克(苏联)。评价：切伦科夫在研究发自镭放射源的辐射穿入不同的液体并被液体吸收时发生的现象中发现的。在切伦科夫之前，也有人观察到当辐射穿入液体时，从液体中会放射出微弱的浅蓝色的辉光，但是它们把它归结为荧光。不过切伦科夫却发现是由入射的辐射引起的。1937 年，弗兰克和塔姆对切伦科夫效应给出了理论解释。他们认为，切伦科夫发现的辐射是由于电子在介质中以大于光在介质中的速度运动时产生的，虽然在真空中无法有速度超过光速(否则与相对论矛盾)，但在介质中光速会减小，$v=c/n$，$n>1$。若 $n<1$，则粒子速度可能会超过光速。就会拖着一条发光的"尾巴"。此外他们是苏联的第一个诺贝尔物理学奖获得者。

**1959 年诺贝尔物理学奖**。发现反质子。获奖者：E. G. 塞格雷、O. 张伯伦(美国)。评价：狄拉克曾预言的反电子已经被发现，人们也相信会有反质子的存在，根据切伦科夫效应制成切伦科夫探测器用以测量高能粒子(毕竟要在介质中超过光速那不可能能量太低)。使高能物理学又有了新的武器。这两位美国人通过新的仪器设备和高超的实验技术证明了反质子的存在。

**1960 年诺贝尔物理学奖**。发现气泡室，取代了威尔孙的云雾室。获奖者：D. A. 格拉塞尔(美国)。评价：这次的科学进步取得得相当偶然。据说一天，格拉塞尔正在喝啤酒，眼睛注视着桌上的啤酒杯，看着那从杯底和杯壁表面不断上冒的一串串气泡，不久，气泡冒完了，他想，冒过气泡的啤酒里也许还会含有气体，他随手将一粒沙子放进啤酒杯中，只见沙子周围又有气泡冒了出来。接着他又放进许多沙子，杯中的啤酒竟像开了锅一样直冒气泡。格拉塞尔从这次"偶然"的发现中得到了启发，要是用液态氢代替啤酒，装入密闭容器，当带电粒子穿过它时，沿途不也会产生一连串的小气泡，从而显示出粒子运动的轨迹来吗？于是，他立即动手试验，终于发明了气泡室。相较于云雾室，气泡室的优点更多，它的空间和时间分辨率高，工作循环周期短，本底干净、径迹清晰，可反复操作。但也有不足之处，那就是扫描和测量时间还嫌太长，体积有限，而且甚为昂贵，不适应现代粒子能量越来越高、作用截面越来越小的要求。不过他还是为物理学家们又提供了一把利刃。

## 思考与练习

1. 三种电场的本质区别是什么？
2. 三种磁场的区别和联系是什么？
3. 谈谈抗磁性的微观解释。
4. 动生电动势与感生电动势有区别吗？为什么？
5. 从狭义相对论的观点，如何理解电场与磁场的相对性？
6. 麦克斯韦为什么能够集电磁理论于大成总结完成经典电磁理论？

第8章

# 光的干涉与时间相干性、空间相干性

光的干涉是光具有波动性的典型特征之一。研究光的干涉的基本方法是进行机械媒质中机械波与电磁以太中电磁波的类比，这是目前国内外大学物理学所采用的研究方法，当前我国高中物理新课程也是按照这个方法。但是，无论是用惠更斯关于经典波动光学的机械以太，还是麦克斯韦关于光的电磁学说的电磁以太，就其物理概念来说都是不正确的。因为迈克耳孙-莫雷(Michelson-Morley)实验和爱因斯坦的狭义相对论彻底地否定了电磁以太的存在，从而在实验和理论上证明了光本身就是一种物质：光波的传播不需要任何介质，光在传播的过程中表现出波动的特性，光在与物质相互作用的过程中表现出粒子的特性，光这种波动性和粒子性完全不同于经典力学里的机械波和质点，现代物理学称之为光的波粒二象性。关于光的波粒二象性的内容在本书第1章和第10章有详细的讨论。

本章将重点讨论相干光的条件、干涉现象中的两个难点内容即时间相干性、空间相干性以及迈克耳孙(A. A. Michelson，1852—1931)干涉仪和法布里-珀罗(Fabry-Perot)干涉仪的基本原理。

## 8.1 相干光的必要条件

干涉是两列(或几列)波叠加时能产生强度在空间呈现稳定的周期性分布的现象，它是波动的最基本特征之一。光的干涉虽然不难实现，但并非任意两光波相遇都能产生干涉现象。从两个完全独立的光源(如两盏电灯)发生的光波即使相遇，我们也难以观察到干涉现象。它们所产生的光强分布等于各单独光强之和，正如同两个人同时唱歌不能产生声的干涉一样。为实现光的干涉现象，相遇的光波必须满足某些条件，我们称这些条件为相干条件，只有当两列波之间满足相干条件时，二者叠加才会形成稳定干涉花样。线性叠加原理是处理光的干涉的理论基础。

### 8.1.1 光波的电磁波特性

光的电磁波特性体现在：其一，电磁波有电场强度矢量 $\boldsymbol{E}$ 和磁场强度矢量 $\boldsymbol{H}$，二者都和传播方向垂直，因此电磁波是横波；其二，对人眼和感光仪器起作用的是电场强度，因此光波的振动矢量指的是电场强度 $\boldsymbol{E}$，把电场强度称为光矢量；其三，在电磁波中，可见光的波长范围为390～760nm，对应的频率范围为 $7.5\times10^{14}\sim4.1\times10^{14}$ Hz。在此范围内，不同的频率引起不同的颜色感觉。

光强即是光波动的能量传递，用平均能流密度来描述。所谓能流密度是指单位时间内通过与波的传播方向垂直的单位面积的能量，而任何波动的平均能流密度与振幅的平方成正比，电磁波的能流密度正比于电场强度振幅 $E_0$ 的平均值，故光强可写为 $\bar{I}\propto E_0^2$。

用相对光强来表征光强而不计算其绝对值,因此可写为 $\bar{I}=E_0^2$。

### 8.1.2 光矢量的叠加

波动叠加的独立性和叠加性原理的特点是:

其一,从几个波源发出的波在同一点相遇时,这些波各自保持自己的波动特性(频率、振幅和振动方向),按照原来的方向传播,彼此互不影响;

其二,线性微分方程的特点:如果有两个独立的函数都满足同一线性微分方程,则这两个函数的线性组合也必然满足这个微分方程,光的波动方程是线性微分方程,因此光波具有叠加性。

干涉现象是波动的特性。若频率相等,在相遇点振动方向几乎沿同一光线,且在观察时间内波动不中断,那么,叠加后产生的合振动在有些地方加强,有些地方减弱,这种现象称为干涉。所以,凡有强弱按一定分布的干涉花样出现的现象都可以作为该现象具有波动性的最可靠最有力的实验依据。

注意,这里仅仅是干涉的角度,讨论电场强度作为光矢量的叠加性原理,更一般地讨论光波的叠加必须运用惠更斯-菲涅耳原理,详细讨论请读者参阅第 9 章。

### 8.1.3 相干必要条件的论证

设有两个点光源 $S_1$ 和 $S_2$,波函数分别为

$$\boldsymbol{E}_1=\boldsymbol{E}_{01}\cos(\omega_1 t-\boldsymbol{k}_1\cdot\boldsymbol{r}_1+\varphi_{01}) \quad (8\text{-}1\text{-}1)$$

$$\boldsymbol{E}_2=\boldsymbol{E}_{02}\cos(\omega_2 t-\boldsymbol{k}_2\cdot\boldsymbol{r}_2+\varphi_{02}) \quad (8\text{-}1\text{-}2)$$

设 $P$ 为两光波叠加区中某一点,如图 8-1-1 所示。应用波的叠加原理,可知 $t$ 时刻,这两列波在空间一点 $P$ 叠加。叠加后的合光矢量可表示为

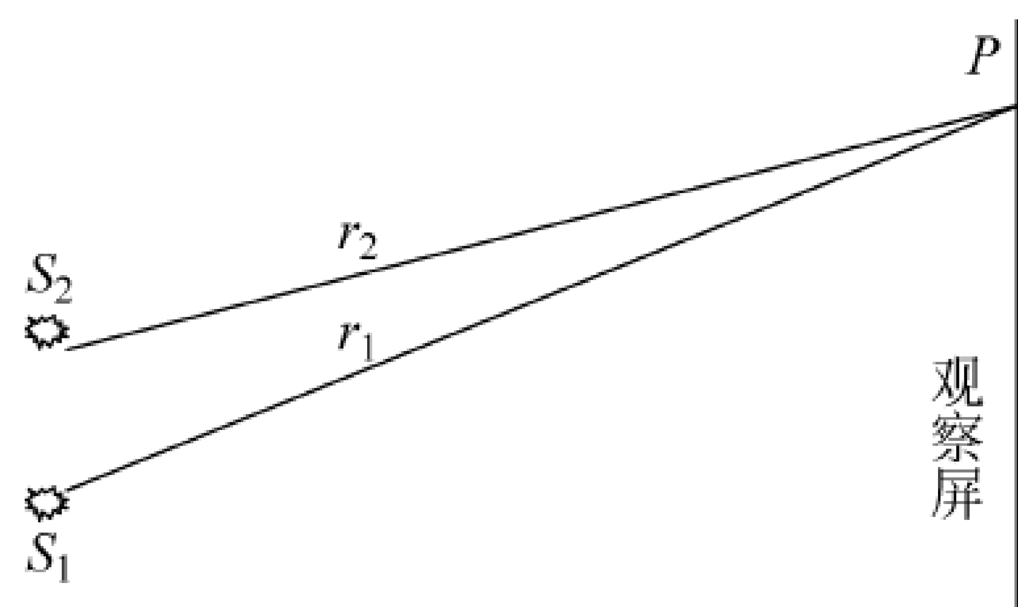


图 8-1-1 讨论相干光必要条件示意图

$$\boldsymbol{E}=\boldsymbol{E}_1+\boldsymbol{E}_2 \quad (8\text{-}1\text{-}3)$$

考察同一种介质中的相对光强,$P$ 点的光强等于合光矢量平方的时间平均值。即

$$\begin{aligned}I&=\langle\boldsymbol{E}^2\rangle=\langle\langle\boldsymbol{E}_1+\boldsymbol{E}_2\rangle\cdot\langle\boldsymbol{E}_1+\boldsymbol{E}_2\rangle\rangle\\&=\langle\boldsymbol{E}_1+\boldsymbol{E}_1\rangle+\langle\boldsymbol{E}_2+\boldsymbol{E}_2\rangle+2\langle\boldsymbol{E}_1\cdot\boldsymbol{E}_2\rangle\\&=I_1+I_2+I_{12}\end{aligned} \quad (8\text{-}1\text{-}4)$$

其中 $I_1=\langle\boldsymbol{E}_1^2\rangle=\frac{1}{2}\boldsymbol{E}_{01}^2$, $I_2=\langle\boldsymbol{E}_2^2\rangle=\frac{1}{2}\boldsymbol{E}_{02}^2$。式中 $I_1$ 表示第一列光波 $E_1$ 单独在 $P$ 点产生的光强;$I_2$ 表示第二列光波 $E_2$ 单独在 $P$ 点产生的光强。所以,按照光干涉的定义,只有当 $I_{12}$ 不为零时,才说明该处发生了光的干涉,因此,称 $I_{12}$ 为两束光干涉的干涉项。不难看出,干涉项的出现是光波叠加的结果。下面重点讨论干涉项。

$$I_{12}=2\langle\boldsymbol{E}_1\cdot\boldsymbol{E}_2\rangle \quad (8\text{-}1\text{-}5)$$

$$\begin{aligned}\boldsymbol{E}_1\cdot\boldsymbol{E}_2&=\boldsymbol{E}_{01}\cdot\boldsymbol{E}_{02}\cos(\omega_1 t-\boldsymbol{k}_1\cdot\boldsymbol{r}_1+\varphi_{01})\times\cos(\omega_2 t-\boldsymbol{k}_2\cdot\boldsymbol{r}_2+\varphi_{02})\\&=\frac{1}{2}E_{10}\cdot E_{20}\{\cos[-\boldsymbol{k}_2\cdot\boldsymbol{r}_2-\boldsymbol{k}_1\cdot\boldsymbol{r}_1+(\omega_2+\omega_1)t+(\varphi_{01}+\varphi_{02})]+\\&\quad\cos[\boldsymbol{k}_2\cdot\boldsymbol{r}_2-\boldsymbol{k}_1\cdot\boldsymbol{r}_1-(\omega_2-\omega_1)t+(\varphi_{01}-\varphi_{02})]\}\end{aligned} \quad (8\text{-}1\text{-}6)$$

在式(8-1-6)中,第一项为和频项,由于对于光波而言 $\omega_1$、$\omega_2\sim10^{15}\,\mathrm{s}^{-1}$,其时间周期 $T=\frac{2\pi}{\omega_1+\omega_2}=2\pi\times10^{-15}\,\mathrm{s}$。而正常眼睛的响应时间为 $\tau=0.1\mathrm{s}$,即使采用高速的探测器($\tau\sim$

$10^{-9}$)，$\cos[-\boldsymbol{k}_2\cdot\boldsymbol{r}_2-\boldsymbol{k}_1\cdot\boldsymbol{r}_1+(\omega_2+\omega_1)t+(\varphi_{01}+\varphi_{02})]$也是一个迅变函数，因此 $T$ 远小于探测器的响应时间 $\tau$，所以，第一项的时间平均值为零。这样，

$$\boldsymbol{E}_1\cdot\boldsymbol{E}_2=\frac{1}{2}\boldsymbol{E}_{10}\cdot\boldsymbol{E}_{20}\cos[\boldsymbol{k}_2\cdot\boldsymbol{r}_2-\boldsymbol{k}_1\cdot\boldsymbol{r}_1-(\omega_2-\omega_1)t+(\varphi_{01}-\varphi_{02})] \quad (8\text{-}1\text{-}7)$$

可见，一般情况下干涉项在空间各点没有确定值，$\cos[\boldsymbol{k}_2\cdot\boldsymbol{r}_2-\boldsymbol{k}_1\cdot\boldsymbol{r}_1-(\omega_2-\omega_1)t+(\varphi_{01}-\varphi_{02})]$在 $\tau_0$ 时间内作随机的迅速变化，多次经历从+1 到−1 之间的一切可能值，导致在观察时间内，$I_{12}$ 的平均值为 0，则 $I=I_1+I_2$，没有干涉现象产生。只有在特定条件下，干涉项的空间各点才有确定值，才会产生干涉现象。

对式(8-1-7)作进一步详细分析。第二项为差频项，只有当时间周期满足 $T_2-T_1=\dfrac{2\pi}{\omega_2-\omega_1}\gg\tau$ 时，其时间平均值才不为零。迄今所知响应最快的探测器的响应时间 $\tau$ 也大于 $10^{-9}$s，这就要求 $\omega_1-\omega_2\ll 2\pi\times10^{-9}\text{s}^{-1}$，才能保证差频项的时间平均值不为零，这个频率差只有 $\omega_1$ 和 $\omega_2$ 的百万分之一。当 $\omega_1$ 和 $\omega_2$ 的差值满足上述条件时，虽然可以探测到由干涉项产生的时间拍频信号，但该信号不能形成稳定的干涉强度的空间分布，只能借助于无线电频率检测或位相检测技术来探测。所以，在光的干涉问题中，为了获得稳定的干涉强度空间分布，它们必须同时满足以下条件。

条件 1：频率相同。即 $\omega_1=\omega_2$ 时，干涉项中的相位差为

$$\Delta\varphi=\boldsymbol{k}_2\cdot\boldsymbol{r}_2-\boldsymbol{k}_1\cdot\boldsymbol{r}_1+(\varphi_{01}-\varphi_{02}) \quad (8\text{-}1\text{-}8)$$

式(8-1-8)表示两光波在空间任一点的相位差只与该点的位置有关，所以，从一点移动到另一点时，随着空间点位置的变化而变化，也就是干涉光强 $I_{12}$ 随空间点位置的变化而变化，从而总光强在空间就出现一定的强度分布。

$\boldsymbol{k}_2\cdot\boldsymbol{r}_2-\boldsymbol{k}_1\cdot\boldsymbol{r}^1$ 对相位差的贡献是由于光从两光源到 $P$ 点所导致的光程差而产生的。$\varphi_{01}-\varphi_{02}$ 则是两光源的初相位差。

条件 2：存在相互平行的振动分量。即在 $P$ 点 $\boldsymbol{E}_{01}$ 和 $\boldsymbol{E}_{02}$ 的振动互相平行或至少不互相垂直，$\boldsymbol{E}_{10}\cdot\boldsymbol{E}_{20}\neq0$。因为若 $\boldsymbol{E}_{01}$ 与 $\boldsymbol{E}_{02}$ 垂直，则 $I_{12}=0$。最理想的条件是 $\boldsymbol{E}_{01}$ 与 $\boldsymbol{E}_{02}$ 平行，这时 $\boldsymbol{r}_1$ 与 $\boldsymbol{r}_2$ 方向一致。由式(8-1-5)、式(8-1-7)、式(8-1-8)得

$$I_{12}=2\langle\boldsymbol{E}_1\cdot\boldsymbol{E}_2\rangle=2\times\frac{1}{2}\boldsymbol{E}_{01}\cdot\boldsymbol{E}_{02}\cos\Delta\varphi=E_{01}E_{02}\cos\theta\cos\Delta\varphi \quad (8\text{-}1\text{-}9)$$

其中 $\theta$ 是振幅矢量 $\boldsymbol{E}_{01}$ 与 $\boldsymbol{E}_{02}$ 的方向在相遇点的夹角。若在 $P$ 点 $\boldsymbol{E}_{01}$ 和 $\boldsymbol{E}_{02}$ 的振动互相平行，$\cos\theta=1$，则，干涉光强就简化为

$$I_{12}=E_{01}E_{02}\cos\Delta\varphi \quad (8\text{-}1\text{-}10)$$

因为 $\boldsymbol{E}_{01}=\sqrt{2I_1}$，$\boldsymbol{E}_{02}=\sqrt{2I_2}$，所以有

$$I_{12}=2\sqrt{I_1I_2}\cos\Delta\varphi \quad (8\text{-}1\text{-}11)$$

若把式(8-1-11)代入式(8-1-10)，则可得光强为

$$I=I_1+I_2\pm2\sqrt{I_1I_2}\cos\Delta\varphi \quad (8\text{-}1\text{-}12)$$

只有两个分量波的振动方向不正交时，才能产生干涉。实际情况经常是，两个分量波的振动方向既不正交，又不平行，这时可将其分解为相互平行和相互垂直振动分量，只有平行分量才能产生干涉。

条件 3：相位差恒定。首先是两光源之间的初相位差 $\varphi_{01}-\varphi_{02}$ 要恒定，然后是两光源到 $P$ 点所导致的光程差 $\boldsymbol{k}_2\cdot\boldsymbol{r}_2-\boldsymbol{k}_1\cdot\boldsymbol{r}^1$ 也要恒定，从而保证 $P$ 点任意时刻的相位差始终

与时间无关,保持不变,即 $\Delta\varphi$ 只是空间位置的函数。式(8-1-8)第二项的时间平均值才不为零。这导致振动的叠加在空间中的分布有稳定的加强和削弱现象,即随空间位置的变化将出现稳定的明暗变化的干涉现象。

$\cos\Delta\varphi$ 取得极大值+1,则 $I$ 达到最大值,称为干涉相长:

$$I_{\max}=(\sqrt{I_1}+\sqrt{I_2})^2 \tag{8-1-13}$$

$\cos\Delta\varphi$ 取得极小值−1,则 $I$ 达到最小值,称为干涉相消:

$$I_{\min}=(\sqrt{I_1}-\sqrt{I_2})^2 \tag{8-1-14}$$

假定在观察时间内,振动时断时续,以致它们的初相位各自独立地作不规则的改变,概率均等地在观察时间内多次历经从 0 到 $2\pi$ 之间的一切可能值,即 $\varphi_{01}-\varphi_{02}=f(t)$,则 $\langle\cos[\boldsymbol{k}_2\cdot\boldsymbol{r}_2-\boldsymbol{k}_1\cdot\boldsymbol{r}_1+(\varphi_{02}-\varphi_{01})]\rangle=0$,即 $I=I_1+I_2$。于是合光强等于分光强之和。

由此可见,光强的叠加可以区分为两种情况:

(1) 振动的相位差在观察时间内无规则地改变(例如间断的振动),合振动的平均强度简单地等于分振动强度之和,不出现干涉现象。通常称这种振动为不相干的。

(2) 两振动的相位差始终保持不变,合振动平均强度可以大于也可以小于分振动强度之和。在这种情况下就可能在较长时间内观察到干涉现象。通常称频率相同、存在相互平行的振动分量且在观察期间内相位差保持不变的两个振动是相干的。

一般来说,参与叠加的各个光波同时满足上述三个条件时,干涉现象一定会发生。所以这三个条件又称为相干的必要条件。满足相干条件的各个光波总的称为相干光波。

下面,对相干必要条件进行进一步的分析。

第一个必要条件,与单色光的要求有关。例如可利用激光器光源(一种准单色光源)并采用适当的分光方法易获得两个以上频率相同的光波。

通过分光方法获得相干光,就要想办法将光源上同一点同一时刻发出的同一列光波“一分为二”,这样分离出来的两列光波由于来自同一波列,频率相同、振动方向相同、初始相位相同,两束光波经不同路径传播后在空间相遇点的相位差恒定,能产生稳定叠加,满足相干条件。分离光波的方法有两种:分波阵面法和分振幅法。

分波阵面法:从波阵面上分离出两部分或更多部分作为初位相相同的相干光源,使各子波源发出的子波在空间经不同路径相遇,产生干涉现象。例如杨氏双缝、洛埃镜、菲涅耳双棱镜以及菲涅耳双面镜等。

分振幅法:利用入射光在薄膜界面上依次反射(或透射)将入射光分解为若干部分(实际上就是将入射光的能量分解为若干部分),经过不同的传播路径再让其相遇,产生干涉现象。例如薄膜(增透膜、增反膜)、劈尖、牛顿环、迈克耳孙干涉仪等。

第二个必要条件,是由于光是矢量波才强调提出的。这一条件易于满足。但为使干涉图样清晰,应尽量使在叠加点引起的振动方向相同。

以图 8-1-2 中的双光束干涉为例。设 $S_1$ 和 $S_2$ 为相同的单色相干点光源,$P$ 为屏上某一叠加点。$L$ 为光源与屏之距离,$d$ 为两点光源之距离。显而易见,当 $d\ll L$ 时,光线 $S_1P$ 与 $S_2P$ 夹角 $\alpha$ 很小,接近平行,因此在 $P$ 点引起的两个振动,可认为其振动方向基本相同。即适当设置实验条件,易于使第二个必要条件满足。

第三个必要条件,涉及干涉场的稳定性问题。对于能够持续发出一定频率的波(如无线电波、声波等)的宏观波源来说,这一条件易于满足。但对于普通光源来说,这一条件不是轻易能够满足的,因为这时微观客体的发光过程主要是自发辐射,它是一种断续的量子过程,

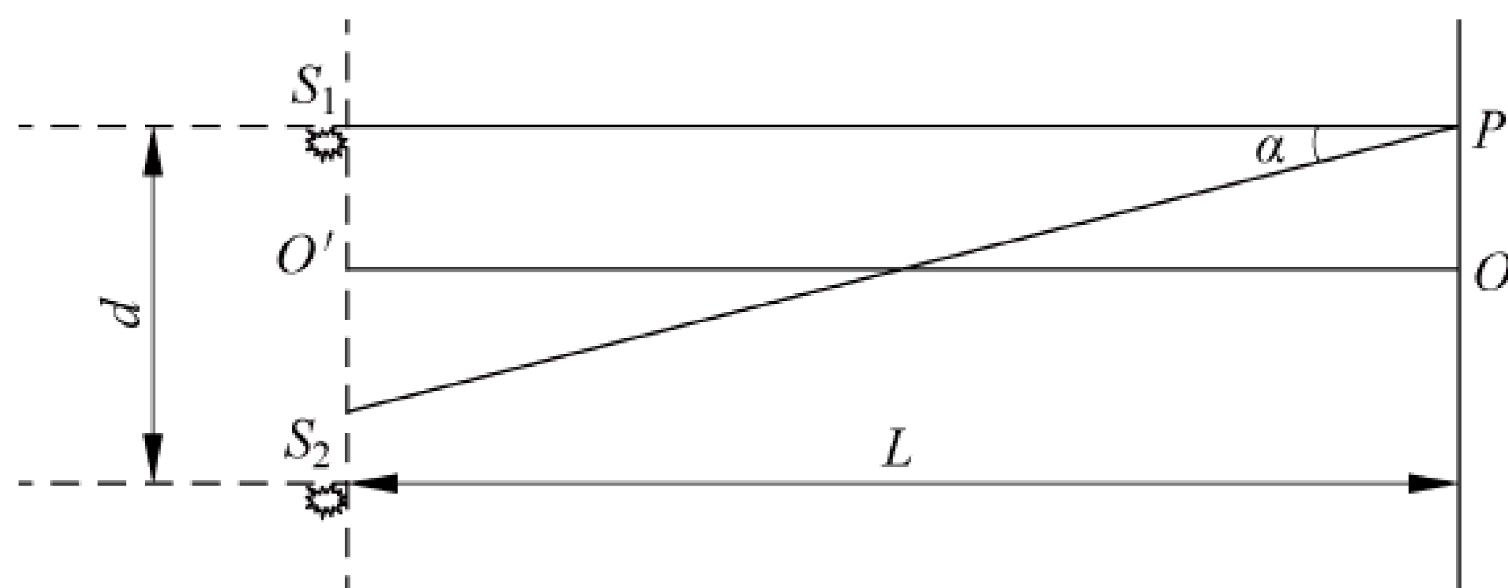



图 8-1-2　双光束干涉

与宏观波源不同；另一方面，由于光波的波长比宏观波源发出的波的波长短许多，因此对实际光源来说，满足第三个必要条件是产生光的干涉的关键。必须设计出相干光源，才能发出相干光波。但是，自从激光器(产生受激辐射)这种实际光源问世以后，易于形成相干光源，从而第三个必要条件的满足也并非难事了。

其实这三个条件并非处于同等地位。第一条是任何波发生干涉的必要条件。第二条是针对矢量波的。一般说来有此二条就足以产生干涉了，剩下的是干涉场的稳定性问题，稳定与否的标准和探测仪器的响应时间有关。对于宏观波源发出的波(如无线电波、声波)，相位差和干涉场的稳定性是不成问题的，相干条件中的第三条无需强调。但对于原子和分子发射的光波来说，这第三条却成了相干条件中最需要着重研究的问题。

对于式(8-1-12)中的干涉项，对于任意两个普通光源(或同一光源的两个不同部分)发出的光波，由于相位差 $\Delta\varphi=\varphi_{01}-\varphi_{02}$ 不固定，$\cos\Delta\varphi$ 的数值在$\pm1$之间迅速地改变着。人们观察到或仪器记录到的是它的时间平均值。在位相的变化完全无规则的情况下，此平均值为0，从而 $I=I_1+I_2$，这时我们说，这两个光源是非相干的，它们的强度是非相干叠加。要产生相干叠加，必须设法使它们发射的光波之间有稳定的相位差。

实现相干光的必要条件，主要途径是将准单色点光源发出的光波分出两束(或多束)光波(主要方法在相干条件一的分析中已给出)，只要它们在叠加区光程差小于波列长度，总是可以相干的。现在的问题是用两个相同的独立光源发出的光波，叠加后能否相干呢？回答这个问题，一方面要看光源本身单色性程度，另一方面还要看光强接收器的时间常数。也就是说要把光源的相干性和观察干涉实验的条件结合起来分析。

通常的低压气体放电管所发出的光波，初相位可保持为常数的时间 $\tau$ 一般不会超过 $10^{-8}\sim10^{-9}$ s，即使用单色性很好的激光光源，初相位保持为常数的时间 $\tau$，目前也只能达到 $10^{-4}$ s 左右。而一般常用的光强接收器，眼睛的时间常数 $\tau_0$ 约为 $10^{-1}$ s，照相器材的响应时间一般不超过 $10^{-3}$ s。显然，只有在接收器的时间常数 $\tau_0$ 范围内光强分布有稳定值，才有可能观测到光的干涉花样。所以，对一般光强接收器(眼睛、照相器材)，$\tau_0\gg\tau$，在 $\tau_0$ 时间内求平均，式(8-2-12)变为 $I=I_1+I_2$，即对一般光强接收器，两个独立光源得不到干涉。但是，近代快速光电接收器件的发展，$\tau_0$ 缩短到 $10^{-9}$ s，有的甚至可到 $10^{-12}$ s，对激光光源 $\tau=10^{-4}$ s 而言，$\tau_0\ll\tau$，$\tau_0$ 在时间内，不仅 $\varphi_{01}-\varphi_{02}$ 几乎没有任何变化，由于$(\omega_{01}-\omega_{02})t$ 项的变化频率$\dfrac{\omega_1-\omega_2}{2\pi}=\Delta\nu=\dfrac{1}{\tau}$，所以$(\omega_1-\omega_2)t$ 项的变化在 $\tau_0$ 时间内也是很小的，即式(8-1-14)中干涉项一般不会是零，出现了干涉现象。事实上也观察到两个独立光源得到的干涉现象。

虽然如此，目前绝大多数干涉装置，主要的还是将同一光源发出的光束分成两束(或多束)，然后使这些光束经过不同的光程后叠加产生干涉。从同一个光束分离出几个光束的方

法一般有分波面法和分振幅法两种。

总之,从以上讨论的各项条件可知:相干条件是对参加叠加的光波的要求,而其着眼点却集中在这些光波对叠加点引起振动的要求。这就是相干条件的实质。

### 8.1.4 相干的补充条件

上述的三个条件是产生干涉现象即相干的必要条件,只满足此条件还不一定就能获得干涉现象,为确保产生明显的干涉现象,还须满足以下两条件:

(1) 两光波在相遇点所产生的振动的振幅相差不悬殊。否则相干叠加的效果不明显,干涉图样不清晰。由式(8-1-12)知,若两光波在相遇点所产生的振幅相差悬殊,例如,$E_{01} \gg E_{02}$,则该点的合成光强 $I$ 将与单一光波在该点所产生的光强 $I_1$ 无实际上的差别,因而观察不出干涉现象。两相干光波的振动方向不垂直只是必要条件,要可见度良好,振动方向应几乎平行为好。

(2) 两光波在相遇点的光程差不能太大。光程差很大时,一光波的波列已通过,而另一光波相应的波列尚未到达,两相应的波列部分重叠,干涉强度较弱,将出现不清晰的干涉现象。换言之,两光波在相遇点的光程差如果大于所用实际光波的相干长度,则无干涉现象。因此,这两个条件可认为是获得干涉现象的补充条件。

### 8.1.5 结论

光相干的三个必要条件是:

(1) 频率相同;

(2) 存在相互平行的振动分量;

(3) 相位差恒定。

两个补充条件是:

(1) 两光波在相遇点所产生的振动的振幅相差不悬殊;

(2) 两光波在相遇点的光程差不能太大。

## 8.2 时间相干性

在第一节讨论光的干涉现象的两个补充条件第二条是两光波在相遇点的光程差不能太大。否则,我们就观察不到稳定的干涉现象。其根本原因是由普通光源的发光特点所决定的。下面作详细讨论。

### 8.2.1 普通光源的发光特点

光是由光源中多个原子、分子等微观客体发射的。微观客体的发光过程是一种量子过程,很难用一个简单的图像描绘清楚。粗略地说,原子或分子每次发射的光波波列都是有限长的,波列的长度与它们所处的环境有关,如果发射光波的原子或分子受到其他原子或分子的作用越强,发射过程受到的干扰越大,波列就越短。不过,即使在非常稀薄的气体中相互作用几乎可以完全忽略的情况下,它们发射的波列持续的时间 $\tau_0$ 也不会大于 $10^{-8}$s,相应的长度小于米的数量级。微观客体的发光过程有两种方式:自发辐射和受激辐射。

**1. 自发辐射的波列长度是有限的**

普通光源(指激光以外的光源)的发射过程以自发辐射为主,这是一种随机过程。光源中原子或分子每一次发光持续时间 $\tau$,即使对持续时间最长的低压气体放电管来说,也不超

过 $10^{-8} \sim 10^{-9}$ s。在这段时间内原子发射波列长度为 $\tau_c$ 的波列，停顿若干时间之后（停顿时间与发光持续时间有相同数量级）再发另一波列，不管是由同一原子或不同原子发出的波列，它们彼此之间在振动方向和相位上都是各自独立的，没有什么联系，没有固定相位差。或者说，最多在发光持续时间（$10^{-8} \sim 10^{-9}$ s）内各波列的位相差才可认为是稳定的。因此，许多继续的波列，持续时间比通常探测仪器的响应时间短得多，振动方向和相位都是无规则的，这就是普通光源发光的基本特征。

激光光源不同于普通光源，它是受激发射占支配地位的光源，激光器中不仅原子发光的持续时间特别长（一般可达到 $10^{-4}$ s）。也就是说，在 $10^{-4}$ s 内初相位可认为是稳定的，而且各发光中心是紧密联系的。也就是说激光的时间相干性和空间相干性都很好。

任何实际光源所发出的单色光都是具有一定波长范围 $\Delta\lambda$ 的准单色光造成谱线宽度的，有四方面的原因。

其一，自然增宽：由能级自然宽度形成。$\Delta\nu = \dfrac{\Delta E_1 + \Delta E_2}{h}$，如图 8-2-1 所示，自然增宽由各能级的寿命决定。

其二，多普勒增宽：分子、原子的热运动引起 $\Delta\nu \propto \bar{v} \propto (T)^{1/2}$，$T \Rightarrow \Delta\nu \uparrow$，即多普勒增宽与温度密切相关。

其三，碰撞增宽：碰撞可增加原子能，$P\uparrow \Rightarrow \Delta\nu\uparrow$，$\Delta\nu \propto \bar{Z} \propto P$ 级宽度，$\Delta E$ 气体 $\Delta\nu \propto \bar{Z} \propto P$（$T$ 一定），碰撞增宽与光源原子或分子之间的相互作用以及温度都有关系。

其四，外界辐射作用引起的增宽：辐射对原子的作用也会增加能级宽度。该作用在微波范围较明显，对可见光的影响不大。

光波中波列是有限长的。普通光源中，原子或分子发射波列所持续的时间为 $\tau_0$，波列的长度为 $Lc = c\tau_0$，称为相干长度。双光束干涉中，要求参与干涉的两列光波到达相遇点的光程差应小于波列的长度，否则，一列波已经通过，另一列波尚未到达而不能发生干涉。可见，波列越长，参与干涉的两列波到达相遇点互相叠加的时间就越长，光场的时间相干性越好。光源发出波长范围为 $\Delta\lambda$ 的准单色光形成的双光束干涉条纹，除了零级条纹外，其他级次条纹之间都要发生相对位移，条纹可见度下降。如图 8-2-2 所示，波长为 $\lambda$ 的 $k$ 级亮条纹的中心位于 $x = k\lambda D/d$，$\lambda$ 的 $k$ 级亮条纹的宽度为 $x = \lambda D/d$。波长范围为 $\Delta\lambda$ 的光的 $k$ 级亮条纹的宽度为

$$\Delta x = k\Delta\lambda \frac{D}{d} \tag{8-2-1}$$

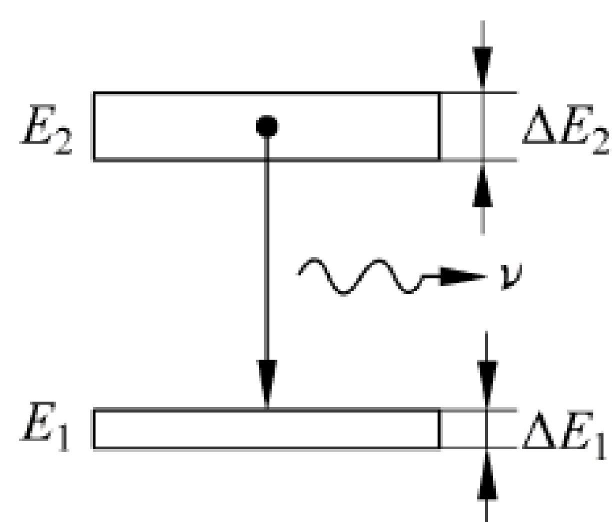



图 8-2-1　能级的自然宽度

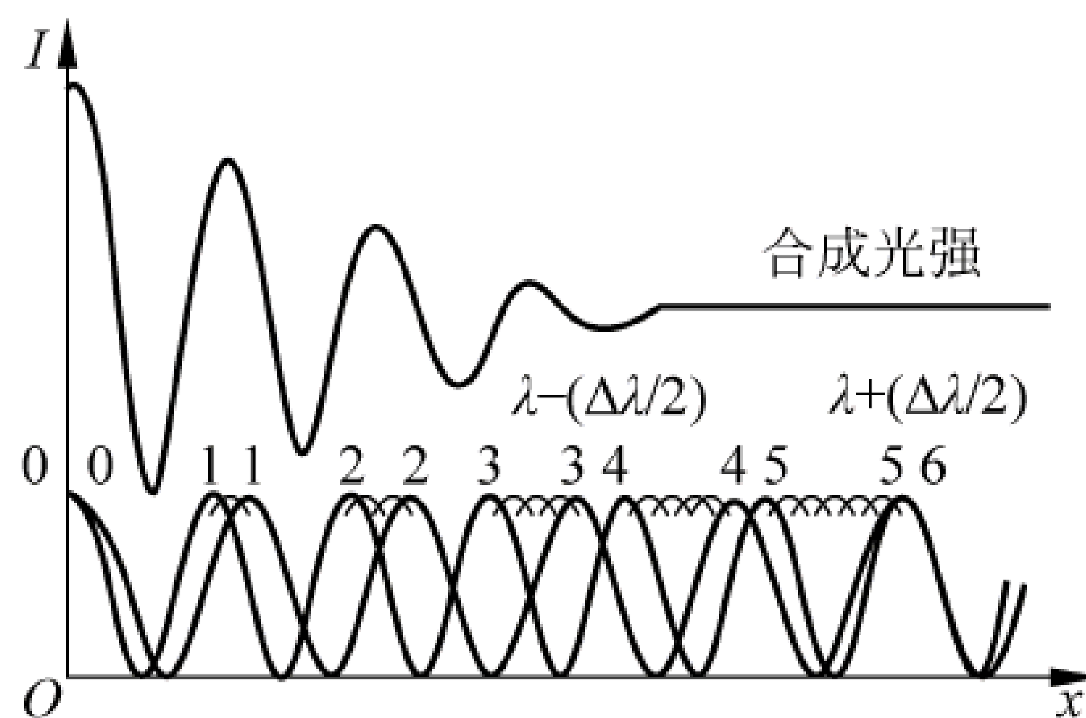



图 8-2-2　光源的非单色性对干涉条纹的影响

由式（8-2-1）可以看出，随着 $k$ 的增加，$\Delta x$ 也相应增大，当 $k_i$ 级亮条纹的宽度正好等于条纹间距时，即 $\Delta x = k\Delta\lambda \dfrac{D}{d} = \lambda D/d$，条纹可见度降为零。于是 $k_i = \lambda/\Delta\lambda$ 与 $k_i$ 级亮条纹

所对应的光程差为

$$\Delta i = k_i(\lambda + \Delta\lambda) = \frac{\lambda(\lambda + \Delta\lambda)}{\Delta\lambda} \approx \frac{\lambda^2}{\Delta\lambda}$$

因此，条纹可见度不为零时，光程差的极限等于波列长度 $L_c$，即 $L_c = c\tau_0 = \frac{\lambda^2}{\Delta\lambda}$。光源的单色性越好，波长范围越小，波列就越长，光的时间相干性就越好。

**2. 最大光程差**

现在来考察一下杨氏干涉实验，如图 8-2-3 所示。

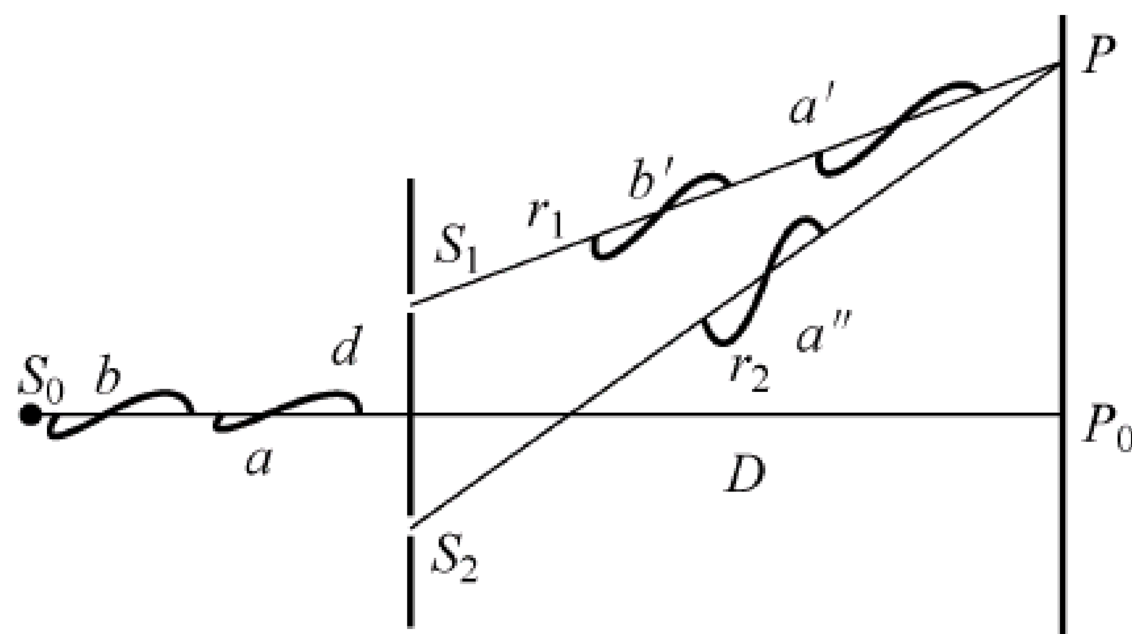



图 8-2-3 杨氏干涉实验示意图

如果光源 $S$ 发射一列光波 $a$，这列光波被杨氏干涉装置分为两个波列 $a'$、$a''$。这两个波列沿不同的路径 $r_1$、$r_2$ 传播后，又重新相遇。由于这两列波是从同一光源发出分割出来的，它们具有完全相同的频率和一定的相位关系，因此可以发生干涉，并可观察到干涉条纹。若两路的光程差太大，致使 $S_1$ 和 $S_2$ 到考察点 $P$ 的光程差大于波列的长度，使得当波列 $a''$刚到达 $P$ 点时，波列 $a'$已经过去了，两列波不能相遇，当然无法发生干涉。而另一发光时刻发出的波列 $b$ 经 $s_1$ 分割后的 $b'$和 $a''$相遇并叠加。但由于波列 $a$ 和 $b$ 无固定的相位关系，因此在考察点 $P$ 无法发生干涉。故干涉的必要条件的是两光波在相遇点的光程差应小于波列的长度。

设光源发出 $\lambda \to \lambda + \Delta\lambda$ 之间的连续波段，由于色散，光束传播越远，各波长的极大值错开越大，当 $\lambda + \Delta\lambda$ 的 $j$ 个极大与 $\lambda$ 的$(j+1)$个极大重合时，其间的各波长极大值均匀错开，使整束光光强为 0，这个距离 $L_c$ 就是这束光的波列长。由 $\delta_{max} = j(\lambda + \Delta\lambda) = (j+1)\lambda$，得到 $j = \frac{\lambda}{\Delta\lambda}$。即得波列的长度 $L_c$ 为

$$\delta_{max} = \frac{\lambda^2}{\Delta\lambda} = \frac{c}{\Delta\nu} = \tau \cdot c = L_c \tag{8-2-2}$$

可见，最大允许光程差 $\delta_{max}$ 就等于相干光的波列长度 $L_c$。也就是说，两列光波要相干涉，其光程差不能超过光波的波列长度。该要求限制了产生干涉的空间区域。这是由于普通光源发光是一种随机过程决定的。否则该部分叠加区域不会满足第三个必要条件。这一点不同于机械波干涉。实际光源所发出的光波，绝非一个无限长的正弦波，而是由一系列有限长的波列组成的。当两光波在相遇点的光程差很小时，则两光波中有固定相位差的波列几乎同时作用于一点，能产生清晰的干涉现象。

## 8.2.2 相干长度和相干时间

**1. 相干长度**

一个波列的长度 $L_c = \tau c$，$\tau$ 为波列持续时间只有同一波列分成的两部分经不同的光程再相遇时才能发生干涉。一旦两光路的光程差 $\Delta L$ 超过波列长度，则不能发生干涉。即

$$\Delta L = L_c = \tau c \tag{8-2-3}$$

最大光程差称作相干长度(coherent length)，很明显它等于一个波列的长度 $L_c$。

**2. 相干时间**

相干时间(coherent time)：光通过相干长度所需时间，也即一个波列的持续时间：

$$\tau = \frac{L_c}{c} = \frac{\Delta L_{\max}}{c} \qquad (8\text{-}2\text{-}4)$$

光源在同一时刻发的光将在不同时刻到达 $P$ 点，若此时间差 $\Delta t = \frac{\Delta L}{c} < \frac{\Delta L_{\max}}{c} = \tau$，还能干涉；或同一时刻到达 $P$ 点的光是光源在不同时刻发出的，若此时间差 $\Delta t < \tau$ 还能干涉。时间相干性讨论的问题就是：$\Delta t$ 为多大就不能发生干涉。很明显，根据式(8-2-2)和式(8-2-4)，$\Delta\lambda \downarrow \Rightarrow \Delta L_{\max} \uparrow \Rightarrow \tau \uparrow$，即时间相干性越好。所以，时间相干性(temporal coherence)是波列长度有限引起的，时间相干性的好坏，就用波列长度 $L_c$——相干长度或波列持续时间 $\tau$ ——相干时间来度量。

### 8.2.3　获得有限波列相干光的两种方法：分割波振面法和分割振幅法

在实验室中通常使用普通光源，同时所使用的绝大部分探测器的光性能响应时间远大于辐射的弛豫时间，因此如何保证两光源的位相差恒定便成为实现干涉的主要矛盾。为了解决发光机制中初相位无规则的迅速变化与干涉要求位相差恒定这一矛盾，基本主导思想是采用自相干的方法，即把同一原子(或点光源)发出的光分解为两列或数列光波(例如利用反射或折射等方法)，然后经过不同的路径使其相遇叠加，这样不管原来光源中各原子发出的光的初位相如何变化，分解出来的各光波到达同一考察点时总保持不变的相位差，所以如此分解出来的光波就成为相干光波。依这种思想分解光波的方法一般有两种：分割波振面法和分割振幅法。

**1. 分割波振面法及实例**

所谓分割波振面法是将一光源的波前(波振面)划分成两个或多个部分，然后再使从它们出射的光相遇叠加，因为它们是从同一光源波前上分割出来的，不管波源的初相位如何变化，它们之间的位相差总保持恒定，因此是相干的，它们相遇叠加时会产生干涉现象。

杨氏(Thomas Young)双缝、菲涅耳(A. J. Fresnel)双棱镜和双面反射镜及洛埃(Floyd)镜等实验都是由波前分割法获得相干光而实现干涉的实验，是光的波动理论坚实的实验基础。虽然扩展光源的照明面积可以较大，但相干面积却很小。为了得到清晰的干涉图样，要求在以上实验中使用的光源前的狭缝在理论上很窄，这是因为普通扩展光源的不同部分是不相干的，否则空间相干性会影响条纹的可见度及条纹数目。

(1) 杨氏双缝干涉

如图 8-2-4 所示，设相干光源 $S_1$、$S_2$ 之间的距离为 $d$，其中点 $M$ 到屏幕 $E$，距离为 $D$，屏幕上任一点 $P$ 到屏幕对称中心的距离为 $x$，点 $P$ 距 $S_1$、$S_2$ 的距离分别为 $r_1$、$r_2$，$PM$ 与 $MO$ 之间的夹角为 $\theta$，由图可见，从 $S_1$、$S_2$ 发出的光到达 $P$ 点的光程差为

$$\delta = r_2 - r_1 = d\sin\theta \approx \frac{d}{D}x \qquad (8\text{-}2\text{-}5)$$

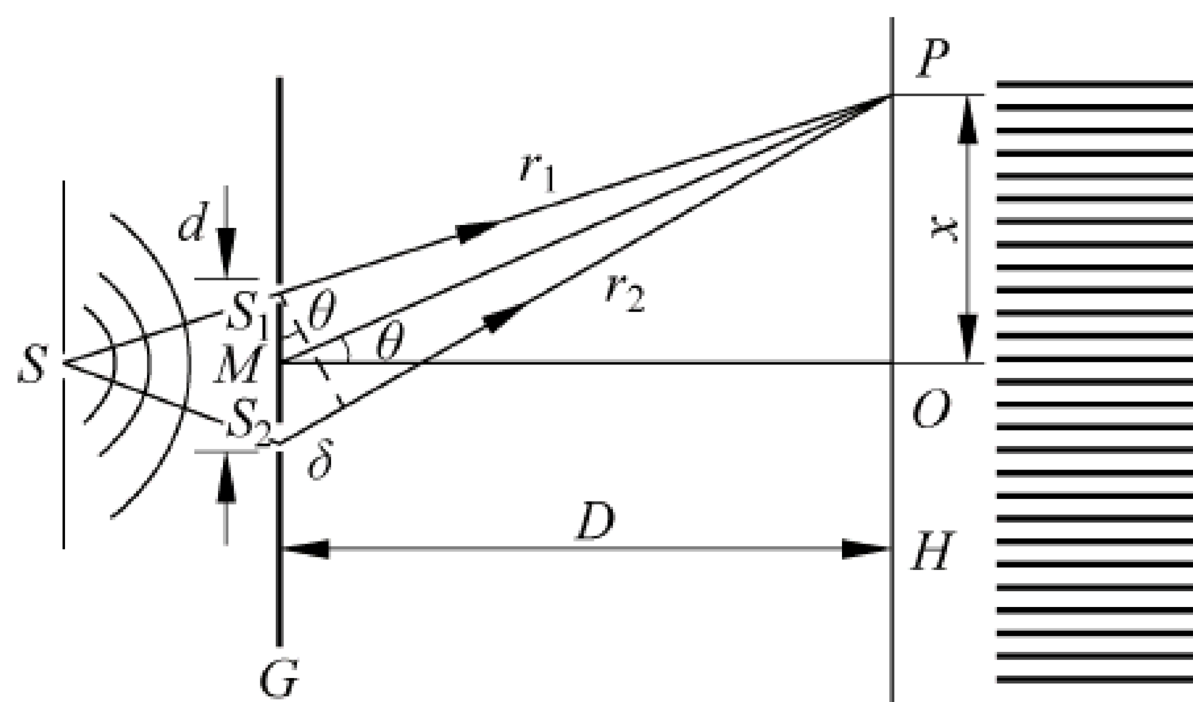



图 8-2-4　杨氏干涉实验示意图

出现明条纹的条件为 $\delta = \frac{d}{D}x = \pm k\lambda, k = 0,1,2,\cdots$。

明条纹中心：$x_k = \pm k\frac{D\lambda}{d}, k = 0,1,2,\cdots$；

出现暗条纹的条件为 $\delta=\dfrac{d}{D}x=\pm(2k+1)\dfrac{\lambda}{2},k=0,1,2,\cdots$。

条纹间距：$\Delta x=x_{k+1}-x_k=\dfrac{D\lambda}{d}$,可见干涉条纹等距对称分布；而且，当 $\lambda$、$D$ 一定时，如果缝宽 $d$ 太大以致 $\Delta x$ 小于可分辨的临界值，将无法观察到清晰的干涉条纹。

对于一定的实验条件，$d$，$D$ 一定，当以白光或复色光入射时，由于波长不同，除中央明纹重叠外，其余各级明暗纹彼此分开，当入射光中 $\lambda_{max}$ 的 $k$ 级明纹与 $\lambda_{min}$ 的 $k+1$ 级明纹重叠时，将无法观察到清晰的干涉条纹。

杨氏双缝干涉条纹的空间分布取决于相干光的光程差 $\delta$，光源的变动(比如单色光波长变化、单色光变为复色光或白光、光源上下移动等)、实验装置的结构变化(比如两缝 $S_1$、$S_2$ 间距变化、宽度变化等，以及光路中放上厚度为 $e$ 的云母片、石英片等)都会引起两束相干光在相遇点的光程差发生变化，从而引起干涉条纹的变动。

(2) 菲涅耳双面镜干涉实验

该实验可以看作为等效的杨氏干涉实验。我们只要求出 $d$ 即可。

图 8-2-5 中屏幕上 $O$ 点在两个虚光源连线的垂直平分线上，屏幕上明暗条纹中心对 $O$ 点的偏离有：

$x=k\lambda\dfrac{D}{d},k=0,\pm1,\pm2\cdots$ 为明条纹中心的位置；

$x=\dfrac{2k+1}{2}\lambda\dfrac{D}{d},k=0,\pm1,\pm2\cdots$ 为暗条纹中心的位置。

(3) 洛埃镜实验

如图 8-2-6 所示，$KP$ 为一背面涂黑的玻璃片，从狭缝 $S_1$ 射出的光，一部分直接射到屏幕 $M$ 上，另一部分经过玻璃片反射后到达屏幕，反射光看成是由虚光源 $S_2$ 发出的，$S_1$、$S_2$ 构成一对相关光源，在屏幕上可以看到明、暗相间的干涉条纹。

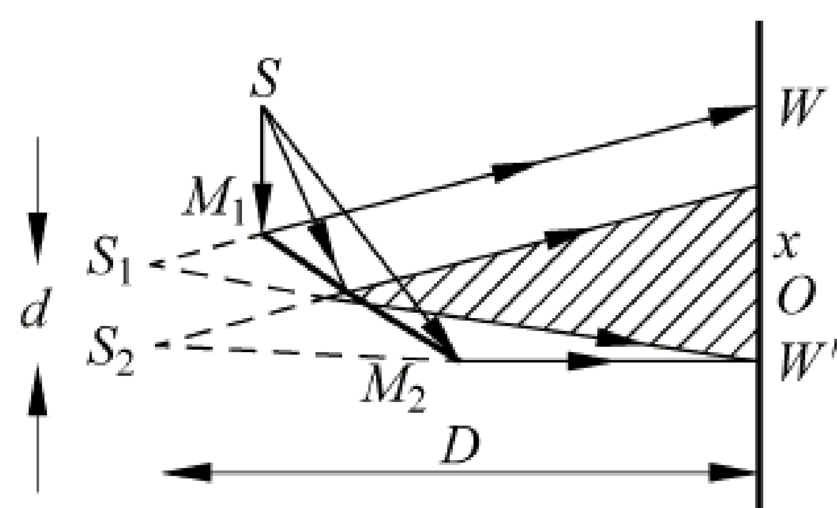



图 8-2-5 菲涅耳双面镜示意图

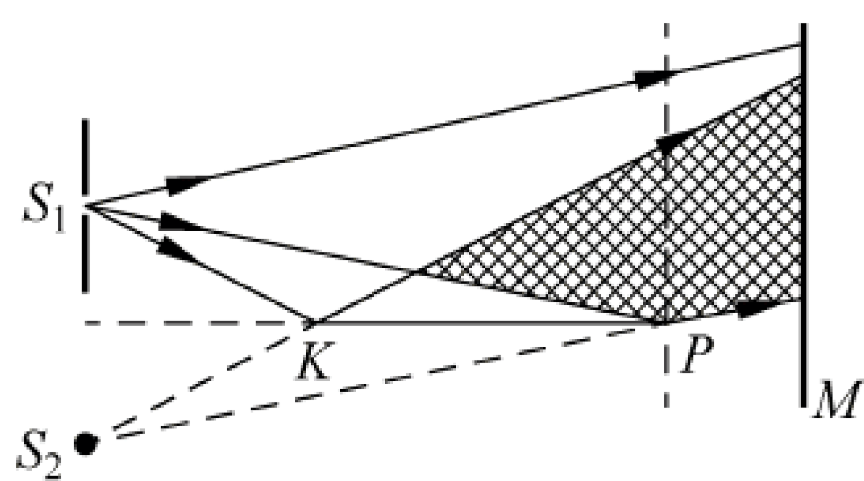



图 8-2-6 洛埃镜示意图

洛埃镜同时也说明了半波损失的存在，当屏幕 $M$ 移至 $P$ 处，从 $S_1$、$S_2$ 到 $P$ 点的光程差为零，但是观察到暗条纹，验证了反射时有半波损失存在。由于半波损失的存在，洛埃镜的明暗纹恰好与杨氏双缝相反。当光从光疏媒质射到光密媒质并在界面上反射时，反射光有半波损失。

**2. 分割振幅法及实例**

当一列光投射在两种透明介质的分界面上时要产生反射和折射，这样一列光波就被分割成为两列光波，即反射光和透射光。因为这两列光波都是源于同一入射光波，入射光波的任何位相改变总是同时发生在这两列光波中，它们之间总保持恒定的位相差，因而它们是相干光波，当设法让它们再次相遇时就会产生干涉现象。正因为它们源于同一光波，它们的振幅和能量也分别取自于同一入射光波的振幅和能量，这种获得相干光波的方法叫作分割振

幅法。

利用分割振幅法产生干涉的典型实验有等倾干涉和等厚干涉，请读者阅读本章第 4 节和第 5 节的详细讨论。

### 8.2.4　小结

综上所述，光的时间相干性和光波的波列的长度密切相关。光源的单色性越好，波长范围越小，波列就越长，光的时间相干性就越好。光的空间相干性与光源线度有关。光源越宽，干涉亮条纹也越宽，当亮条纹宽度等于条纹间距时，干涉条纹的可见度降为零。在杨氏实验中，为改善干涉条纹可见度，实验中我们通常在光源前放置狭缝以减小光源的线度，提高光场的时间相干性。另外由于激光的时间相干性较好，故将激光直接投射在双缝上也可以获得良好的可见度。

## 8.3　空间相干性

所谓空间相干性就是对一般的普通光源（所谓的非相干光源），通过对干涉装置的空间几何参数控制，使其能够实现干涉的最小几何线度。

### 8.3.1　光源线度 $b$ 对干涉条纹影响

比如在杨氏双缝干涉实验中，设普通光源前有一个单狭缝 $S$，其开缝宽度为 $b$，在 $S$ 的正前方 $l$ 处对称地设置两个狭缝 $S_1$ 和 $S_2$，使 $S_1$ 与 $S_2$ 之距为 $d$，如图 8-3-1 所示。则在理想线光源情况下（即认为 $b$ 很小，或 $b\to 0$；暂不考虑光源的时间相干性，即认为光源是单色的长波列光源），其干涉条纹的间距（明暗条纹宽度之和）为 $E=\dfrac{L}{d}\lambda$。由于实际光源 $\Sigma$ 有一定扩展面积，即缝 $S$ 有一定的宽度 $b$，致使干涉条纹的间距加宽，增加的宽度为 $\Delta X=2\Delta X_{MN}=b\dfrac{L}{l}$。

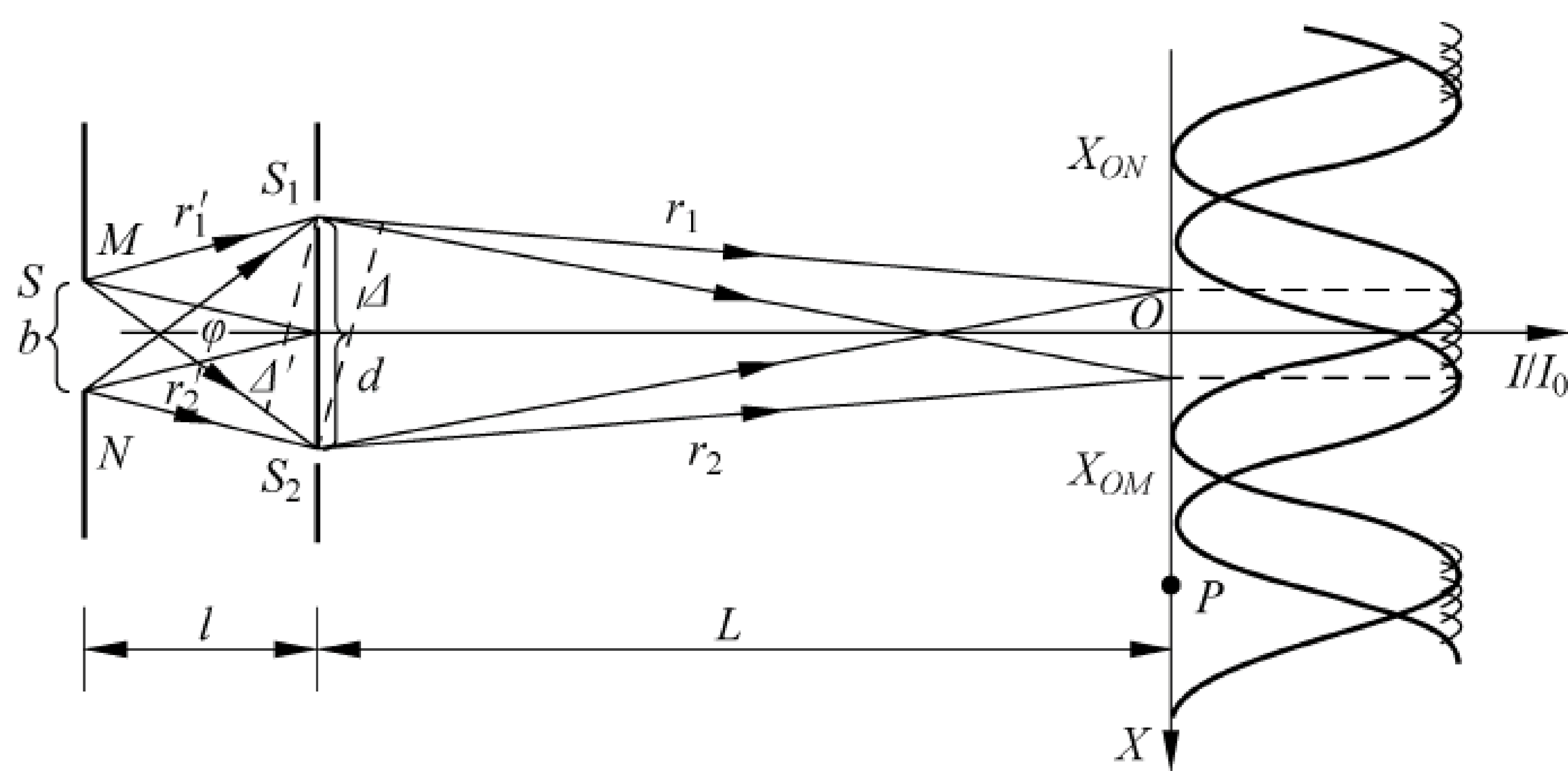



图 8-3-1　空间相干性干涉示意图

当干涉装置的几何尺寸 $L$、$d$、$b$ 和 $l$ 相互变化，使条纹增加的宽度恰等于当 $b\to 0$ 时条纹的间距时，干涉花样消失，即 $\Delta X=E$，$b\dfrac{L}{l}=\dfrac{L}{d}\lambda$ 时，得 $b_e=\dfrac{l_e}{d_e}\lambda$，条纹的视见度 $V=0$（$I_{\max}=I_{\min}$）。此时，相当于在实际光源 $\Sigma$ 前去掉了 $S$ 狭缝的约束，即实际装置的参数 $b_e$、$l_e$ 和 $d_e$ 决定了该普通非相干光源能够产生干涉的临界尺寸线度。也就是说，此时如果任

一非相干光源 Σ 的尺寸 $b<b_e=\frac{l_e}{d_e}\lambda$ 而不需加任何约束，其前面的 $S_1S_2=d_e$ 且距光源 Σ 有 $l_e$ 之距，就会实现干涉。如果其线度 $b\geqslant d_e$ 且 $d\geqslant d_e$，就不会产生干涉。

如图 8-3-1 所示，由于光源线度扩展为 $b$，其上 $M$ 点发出的光到达屏上 $P$ 点的光程差为

$$\Delta_{MP}=(r_1+r_1')-(r_2+r_2')=(r_1'-r_2')+(r_1-r_2)=+\Delta'+\Delta \tag{8-3-1}$$

$l\gg d$，且 $l\gg b$ 时，有 $r_1'-r_2'\approx-\frac{bd}{2l}$，$r_1-r_2\approx d\frac{x}{L}$，则

$$\Delta_{MP}=-\frac{bd}{2l}+d\frac{x}{L} \tag{8-3-1$'$}$$

因此，由光源 $S$ 上(而不是 Σ 上)的 $M$ 点产生的各级亮条纹的位置如下：

$$d_e\ \text{零级亮纹：}\Delta_{MP}=0\ \text{得}\ X_{OM}=+\frac{bL}{2l} \tag{8-3-2}$$

$$k\ \text{级亮纹：}\Delta_{MP}=k\lambda\ \text{得}\ X_{KM}=\frac{L}{d}k\lambda+\frac{bL}{2l} \tag{8-3-3}$$

同理，$S$ 上 $N$ 点产生各级亮条纹位置由如下的光程差决定：

$$\Delta_{NP}=(r_1+r_1')-(r_2+r_2')=(r_1'-r_2')+(r_1-r_2)=+\Delta'+\Delta \tag{8-3-4}$$

同样的几何条件有：$r_1'-r_2'=+\frac{bd}{2l}$，$r_1-r_2=d\frac{x}{L}$，则

$$\Delta_{NP}=\frac{bd}{2l}+d\frac{x}{L} \tag{8-3-4$'$}$$

$$\text{零级亮条纹}\quad \Delta_{NP}=0\quad \text{得}\quad X_{ON}=-\frac{bL}{2l} \tag{8-3-5}$$

$$k\ \text{级亮条纹}\quad \Delta_{NP}=k\lambda\quad \text{得}\quad X_{KN}=\frac{L}{d}k\lambda-\frac{bL}{2l} \tag{8-3-6}$$

其结果是：原来 $b\to0$ 时各级条纹的位置，由于 $b$ 的增宽，向其两边各移动了 $\frac{bL}{2l}$。$M$ 源点使条纹下移(向 $X$ 轴正向)即 $+\frac{bL}{2l}$，就中心零级条纹而言移至 $X_{OM}$ 处，如图 8-3-1 所示；同理，$N$ 源点使其上移($X$ 轴负向)，即 $-\frac{bL}{2l}$，零级移至 $X_{ON}$ 处。

(1) 非相干光源线度的 $b<b_1=\frac{l}{4d}\lambda$，零级亮条纹的变动

此时如图 8-3-2 所示，零级亮纹由于源点 $M$ 和 $N$ 的影响而上下移至 $X_{OM}$ 和 $X_{ON}$ 处，即由 $O$ 点向两边各扩展了 $\frac{bL}{2l}$ 而使错开的间隔为 $\Delta X_1=2X_{OM}=2X_{ON}=X_{ON}-X_{OM}=\frac{Lb}{l}$。

因为 $b<b_1=\frac{l}{4d}\lambda$，所以 $\Delta X_1$ 小于条纹间距的四分之一，即 $\Delta X_1=\frac{Lb}{l}<\frac{Lb_1}{l}=\frac{l}{4d}\lambda$

其他级次条纹的移动也如此，干涉花样中 $I_{\max}$ 与 $I_{\min}$ 相差较大，因而视见度较好。

(2) 非相干光源线度 $b=b_2$，且 $\frac{l}{4d}\lambda<b_2<\frac{l}{2d}\lambda$

这种情况下，条纹错开的间隔介于在 $b\to0$ 时条纹间距的四分之一与二分之一之间，即

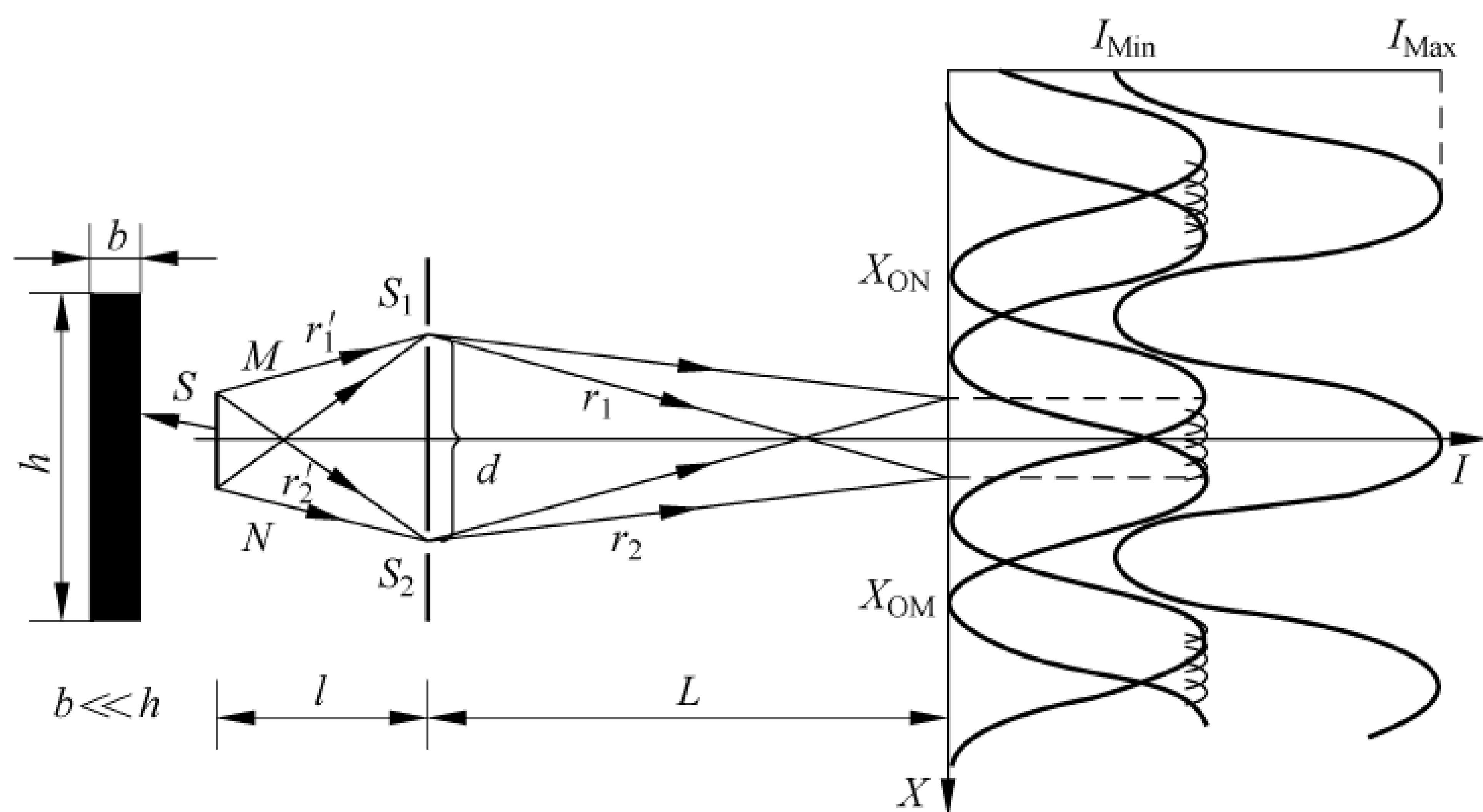



图 8-3-2　零级亮条纹的变动

条纹移动的距离为 $\Delta X_2=2X_{OM}=2X_{ON}=\dfrac{Lb_2}{l}$，所以$\dfrac{l}{4d}\lambda<\Delta X_2<\dfrac{l}{2d}\lambda$ 的干涉花样中，$I_{\max}$ 与 $I_{\min}$ 之差变小，视见度大为下降，如图 8-3-3 所示。

(3) 非相干光源宽度的 $b=b_3=\dfrac{l}{d}\lambda$

这时，条纹错开的间隔恰等于 $b\to0$ 时条纹间距，即 $\Delta X_3=2X_{OM}=2X_{ON}=\dfrac{Lb_3}{l}$，亦即 $\Delta X_3=\dfrac{L}{d}\lambda$ 时的干涉花样中，$I_{\max}=I_{\min}$，屏上一片明亮，视见度为零，如图 8-3-4 所示。

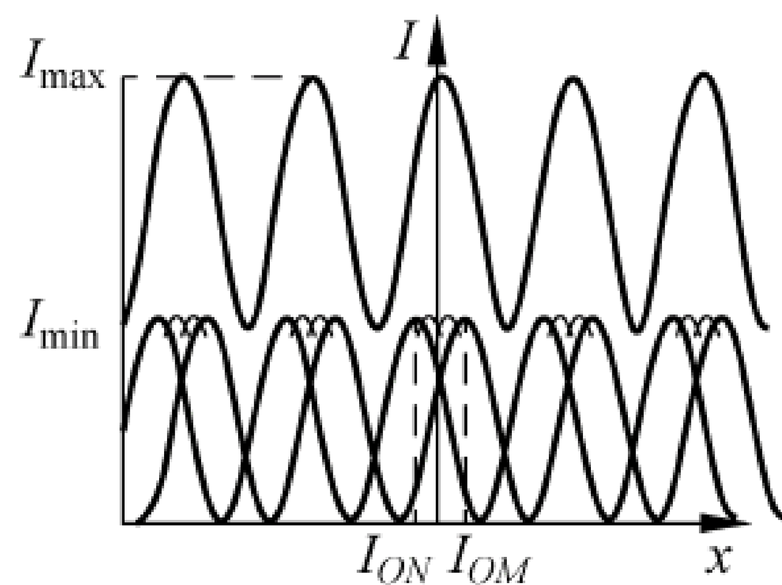



图 8-3-3　视见度大为下降

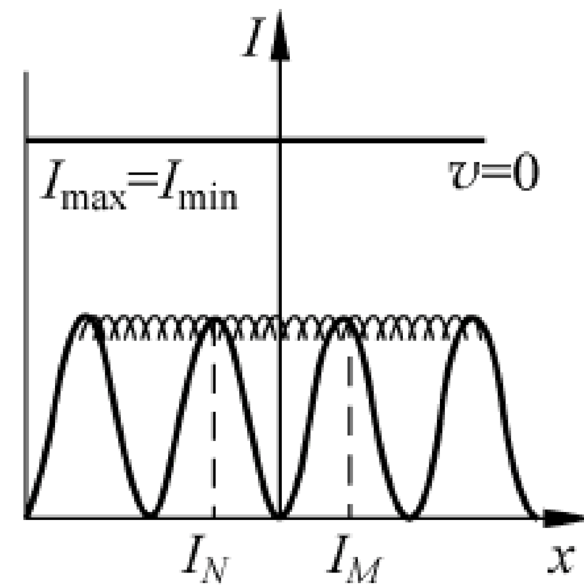



图 8-3-4　视见度下降为零

## 8.3.2　*b* 一定，*L* 和 *d* 对干涉条纹影响分析

当理想线光源($b\to0$)的条纹间隔$\dfrac{L}{d}\lambda$ 等于扩展非相干光源($b>0$)时条纹错开移动的间隔$\left(\Delta X=\dfrac{Lb}{l}\right)$时，即$\dfrac{L}{d}\lambda=\dfrac{Lb}{l}$，可得 $d_e=\dfrac{l_e}{b_e}\lambda$，则称 $d_e$ 为扩展非相干光源在宽度为 $b_e>L_e$ 和 $b_e$ 条件下的相干间距。干涉装置在相干间距内变化时，条纹移动错开会影响干涉花样的视见度，超过 $d_e$ 所界定的参数，则其不能实现干涉。

**1. *L* 不变，*d* 变化的情形**

参见图 8-3-1，当 $d$ 增加时，由于干涉条纹移动的间隔 $\Delta X_{MN}=\dfrac{Lb}{l}$与 $d$ 无关而不变，但条纹间距 $E=\dfrac{L}{d}\lambda$($b\to0$ 时)而变窄，因而很容易造成 $\Delta X_{MN}>E$，所以 $d$ 增加的过程，也就

是等效于 $\Delta X_{MN}$ 相对增加的过程,干涉花样中的 $I_{\max}$ 与 $I_{\min}$ 之差逐渐变小,视见度逐渐下降,如图 8-3-5(a)和(b)所示。$d$ 的变化必须满足 $d \ll d_e = \frac{l_e}{b_e}\lambda$ 时,才会使 $I_{\max}$ 与 $I_{\min}$ 有较大的差值而保证有较好的视见度。当 $d \gg d_e$ 时 $\Delta X_{MN} \geqslant E$,而使其视见度不会为零。

同理,当 $d$ 减小时,条纹间距 $E$ 变宽,由于 $\Delta X_{MN}$ 不变,则 $d$ 减小的过程也就是等效 $\Delta X_{MN}$ 相对减小的过程,因而不易造成 $\Delta X_{MN} > E$,所以,视见度 $V$ 会增加。$d$ 越小,干涉花样清晰。

**2. *d* 不变,*l* 变化的情形**

当 $l$ 增加时,由于 $E = \frac{L}{d}\lambda$ 与 $l$ 无关而不变,其条纹移动间隔 $\Delta X_{MN} = \frac{Lb}{l}$ 而减小,这样 $E > \Delta X_{MN}$ 容易满足,因此干涉花样的视见度 $V$ 会增大。如图 8-3-6 所示。另一方面,由 $d_e = \frac{\lambda}{b/l} = \frac{\lambda}{\varphi}$,即 $\varphi = \frac{b}{l}$ 参见图 8-3-1。当 $l$ 增加时,$\varphi$ 减小,则 $d_e$ 增加,即整个干涉装置的相干间距增加,因为 $\varphi$ 减小意味着光源 $S$ 远离 $S_1$ 和 $S_2$,因此:$d_e = \frac{\lambda}{\varphi}$ 揭示了足够远的、线度为 $b$ 的非相干光源也有一定的相干间隔,这就从光源的空间方面揭示了相干光源空间横向最大两点在同一时刻的关联程度,即 $l$ 适当增加并控制 $b$、$d$ 等参数,干涉就可以实现。

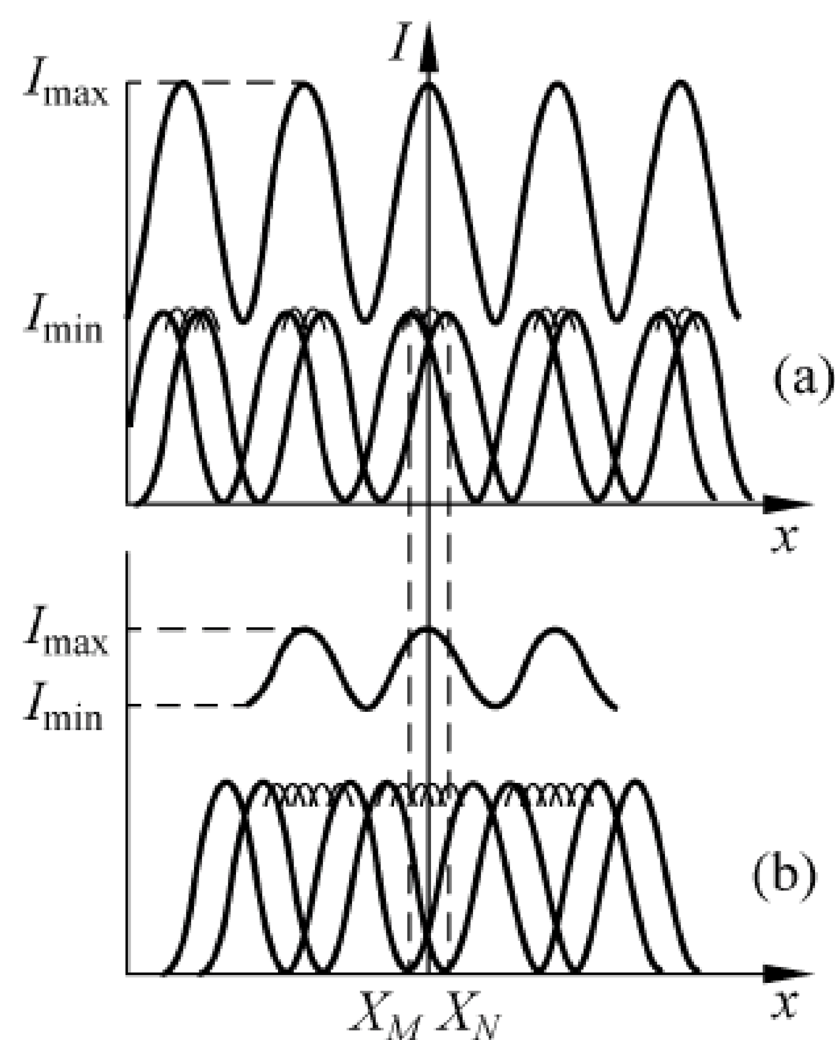


图 8-3-5 视见度逐渐下降

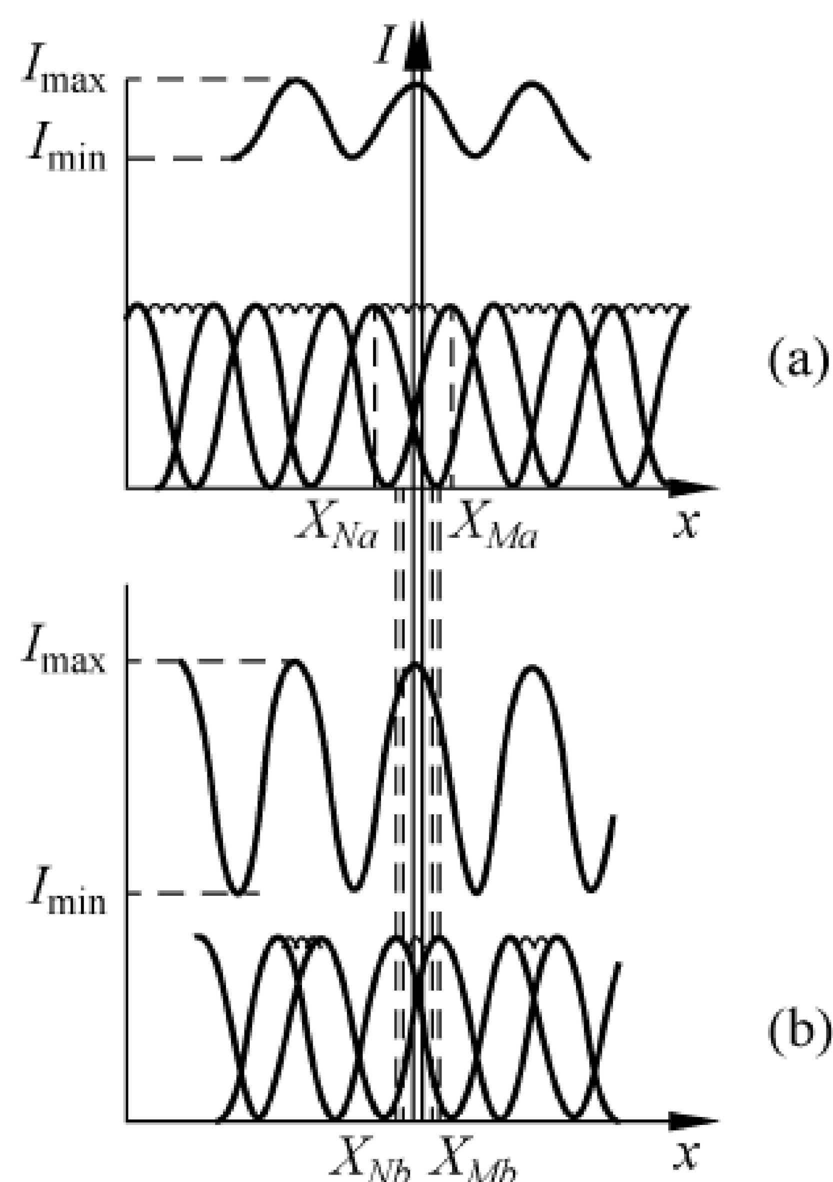


图 8-3-6 视见度逐渐增加

同理,当 $l$ 减小时,$E = \frac{L}{d}\lambda$ 仍不变,$\Delta X_{MN} = \frac{Lb}{l}$ 而增加,容易造成 $\Delta X_{MN} > E$,使条纹视见度 $V$ 下降到零;同时,$d_e = \frac{\lambda}{\varphi}$,$l$ 减小,$\varphi$ 增加而使 $d_e$ 减小,此意味着非相干普通光源的横向线度足够大,以至于其上横向最大两点的光波场在同一时刻不满足相干条件,使干涉不能实现;从实验上考虑,由于 $d_e$ 很小,也难以实现 $d \ll d_e$。

## 8.3.3 结论

任何非相干光源,都有可能实现其干涉。其横向线度 $b$ 与实验装置的其他参数必须满足一定的条件,形成特定的制约关系。就分波前双光束装置而言,从视见度 $V$ 考虑,其临界

条件是$\frac{L}{d}\lambda=\frac{Lb}{l}$，由此形成的制约关系要求$b<\frac{1}{4}b_e=\frac{\lambda l}{4d}$和$d\ll d_e=\frac{l}{b}\lambda$，这就是所谓的空间相干性，其物理本质是非相干普通光源所发随机波列光波场在空间尺度上关联程度的宏观反映。

光的空间相干性和时间相干性是不能严格分开的，例如杨氏实验中，考察屏幕上离$P$点较远位置处的干涉条纹时，由于光波分别从$S_1$和$S_2$传播到这些点所需的时间不同，因此有较大的光程差。由此可见，空间相干性和时间相干性问题是密切联系的，不能绝对分开。但是这并不代表它们之间没有区别。从上面的讨论，光的干涉中的时空相干性的本质在于光既具有波动性又具有粒子性。但是，空间相干性和时间相干性是两个不同的概念，虽然它们都是看光场中是否会产生相干的问题，然而，空间相干是由面光源的不同部分发光的独立性引起的，表现在光场的横方向（波前）上，一般属于分波前干涉。而时间相干性是由于光源发光过程在时间上的不连续性引起的，表现在光场的纵方向（波线）上，一般属于分振幅干涉。

综合前面的讨论可知，影响干涉条纹的视见度共有六个因素：三个必要条件；两相干光束的振幅比；光源的非单色性（时间相干性）和光源的尺寸大小（空间相干性）。实际情况往往是多个因素共同作用。当从理论上确保了产生干涉条纹可见度较好的条件后，为了能用眼睛直接观察清楚，条纹间距也不能太小。由于两个相邻级次的亮纹之间的光程差为一个波长，一般光波的波长很小，所以实验时还必须精心布置干涉装置，以保证干涉条纹的分布有足够的间距。

## 8.4　迈克耳孙干涉仪

干涉仪是利用干涉条纹的变动来测量长度或介质折射率微小改变的仪器。迈克耳孙干涉仪是迈克耳孙（A. A. Michelson，1852—1931）在1881年根据分振幅干涉原理设计成功的。最早设计出的这种干涉仪构思精巧，结构简单，有着广泛的用途。迈克耳孙和莫雷（Edward Williams Morley，1838—1923）应用该仪器进行了测定以太风的著名实验，后人根据此种干涉仪研制出各种具有实用价值的干涉仪。

利用迈克耳孙干涉仪可观察单色光或复色光的等倾干涉条纹（明暗相间的同心圆）和等厚干涉条纹（明暗相间的平行线），测量单色光的波长、复色光的波长平均值，亦可测双色光的波长差，迈克耳孙干涉仪是光学实验主要设备之一。

### 8.4.1　迈克耳孙干涉仪的结构和基本原理

图8-4-1是迈克耳孙干涉仪的光路图。图中$M_1$、$M_2$是两面精密磨光的平面反射镜。$M_1$的位置是可以移动的，借助一微动的螺旋丝杆，可以在图中所示的箭头方向平移。大型的干涉仪，$M_1$的移动范围可达几十厘米。$M_2$的位置是固定的。利用$M_1$、$M_2$上的螺旋，可以调节$M_1$、$M_2$两镜面互相垂直，或有微小夹角。$S$是面光源，例如钠光灯或扩束的激光等，有时为了扩大光源的面积，可在光源前插入一毛玻璃或透镜等。$E$是接收干涉纹的器件，例如，人眼或望远镜、照相机等。$G_1$、$G_2$是两块折射率和厚度都完全相同的

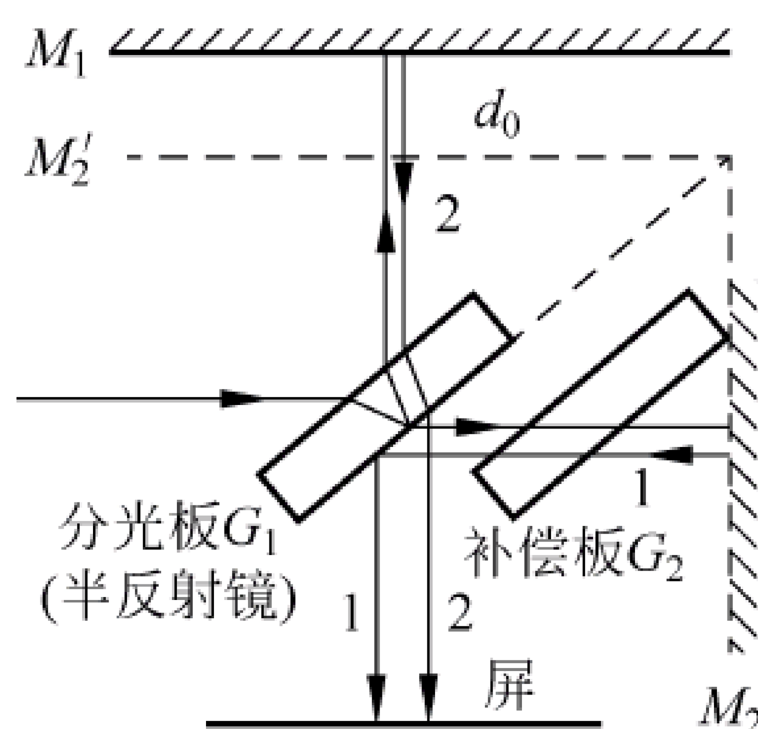


图8-4-1　迈克耳孙干涉仪的光路图

光学平行平面玻璃板，两者必须彼此严格地平行(在制造时，先将一整块玻璃板磨成两面严格平行的光学平面，然后将它分割成完全相同的两块)。$G_1$ 的法线方向与 $M_1$ 移动方向成45°角，在 $G_1$ 板的一面镀有一层半反射膜，可使透射的光，一半反射一半透过，故称分光板。$G_2$ 称补偿板，其与 $G_1$ 平行。经 $M_2$ 反射的光三次穿过分光板，而经 $M_1$ 反射的光只通过分光板一次，补偿板就是为了消除这种不对称而设置的。在使用单色光源时，补偿板并非必要，可以利用空气光程来补偿；但在使用复色光源时，因玻璃和空气的色散不同，补偿板则是不可缺少的。若要观察白光的干涉条纹，两相干光的光程差要非常小，即两臂基本上完全对称，此时可以看到彩色条纹；若 $M_1$ 或 $M_2$ 稍作倾斜，则可以得到等厚的交线处($d=0$)的干涉条纹为中心对称彩色直条纹，中央条纹由于半波损失为暗条纹。

从单色光源 $S$ 发出的光线射至分光板 $G_1$，一半反射至 $M_1$，经 $M_1$ 反射，再透过 $G_1$ 射向接收器 $E$(透射途中被 $G_1$ 的半透明表面所反射的部分散射到空中，故不作讨论)；另一半透过 $G_1$、$G_2$，射向 $M_2$，再相继反射，也射向 $E$(反射途中穿过 $G_1$ 半透明表面的部分也散至空中，不予讨论)。这样由一束光分成的两束光是相干的。在 $E$ 处干涉情况，决定于光线自分光板 $G_1$ 分裂为二以后，各自经历的光程之差。为了便于考虑这样两支相干光线到达 $E$ 处的光程差，可以考虑 $M_1$ 经分光板 $G_1$ 成虚像于 $M_1'$。由于平面镜成像的对称性，光线自 $G_1$ 到 $M_1$ 来回的光程，将等于自 $G_1$ 到 $M_1'$ 来回的光程。于是相干两光线到达 $E$ 处的光程差，就可以看成光线经 $M_2$、$M_1'$ 之间的"空气膜"两表面反射的光程差，也就相当于薄膜干涉了。

### 8.4.2 干涉条纹

当调节 $M_1$、$M_2$，使之互相垂直，即 $M_2$、$M_1'$ 互相平行。观察者的眼睛或望远镜对无穷远调焦时，就可以看到圆环形的等倾干涉条纹。若 $M_1$ 和 $M_2$ 并不恰好垂直，即 $M_2$、$M_1'$ 有微小的夹角，就会在它们的表面看到楔形"空气层"的等厚条纹。那时，眼睛或望远镜将对"空气层"表面调焦。

现在来讨论当 $M_1$ 移动时，干涉条纹变化的情况。

(1) 等倾干涉

设 $M_1$ 和 $M_2$ 垂直(即 $M_1'$ 与 $M_2$ 平行)，并且 $M_1$ 至 $G_1$ 距离小于 $M_2$ 至 $G_1$ 的距离(即 $M_1'$ 在图 8-4-1 中 $M_2$ 的左边)，这时在 $E$ 处看到圆环等倾干涉条纹。由薄膜干涉的光程差公式 $\delta=2d\sqrt{n_2^2-n_1^2\sin^2 i_1}-\dfrac{\lambda}{2}$ 或 $\delta=2dn_2\cos i_2-\dfrac{\lambda}{2}$ 可知，$n_1=n_2=n=1$，不发生折射，$i_1=i_2$，且没有额外光程差。令 $i_1=i_2=i'$，则可得

$$\delta=2d\cos i'\begin{cases}k\lambda \rightarrow \text{相长}\\ (2k+1)\dfrac{\lambda}{2} \rightarrow \text{相消}\end{cases} \tag{8-4-1}$$

其中，$d$ 为"空气薄膜"的"厚度"，$i'$ 为光在平面镜 $M_1$ 上的入射角。当 $M_1$ 逐渐远离 $G_1$ 时，圆环形干涉纹中心，将明暗交替地变化。由式(8-4-1)可得，在干涉纹的中心点，即 $i'=0$，故

$$\delta=2d\begin{cases}k\lambda \rightarrow \text{相长}\\ (2k+1)\dfrac{\lambda}{2} \rightarrow \text{相消}\end{cases} \tag{8-4-2}$$

可见，当 $M_1$ 在某一位置时，中心点恰好最亮，则 $M_1$ 每移动 $\lambda/2$ 的距离($d$ 改变 $\lambda/2$)，

即光程差 $\delta$ 改变了一个波长，故中心点将由暗再到亮，经历了一次亮暗的变化。若 $M_1$ 继续移动，中心点将亮、暗周期地变化。中心点光强变亮（或变暗）的次数，也就是 $M_1$ 移动了多少个 $\frac{\lambda}{2}$。

再讨论当 $M_1$ 移动时，干涉图样中心以外的环形条纹变化的情况。由

$$2d\cos i' = k\lambda \tag{8-4-3}$$

可知，当 $d$ 一定时，中心点（$i'=0$）$k$ 有最大值。$i'$ 愈大，$k$ 愈小，即半径愈大的环形干涉纹，干涉级次愈低。当我们看着某一亮纹（即 $k$ 为一定值），若 $M_2$ 移动 $d$ 减小时，由式(8-4-3)可知，$i'$ 也必须减小（才能保持 $k\lambda$ 为定值）。这样，我们将看到该亮纹向中心移动。$M_1$ 每移动 $\frac{\lambda}{2}$ 的距离（即 $d$ 减少 $\lambda/2$），将有一个亮纹收缩到中心点，同时，条纹之间的距离也逐渐增大（环纹变稀），如图 8-4-2 所示。

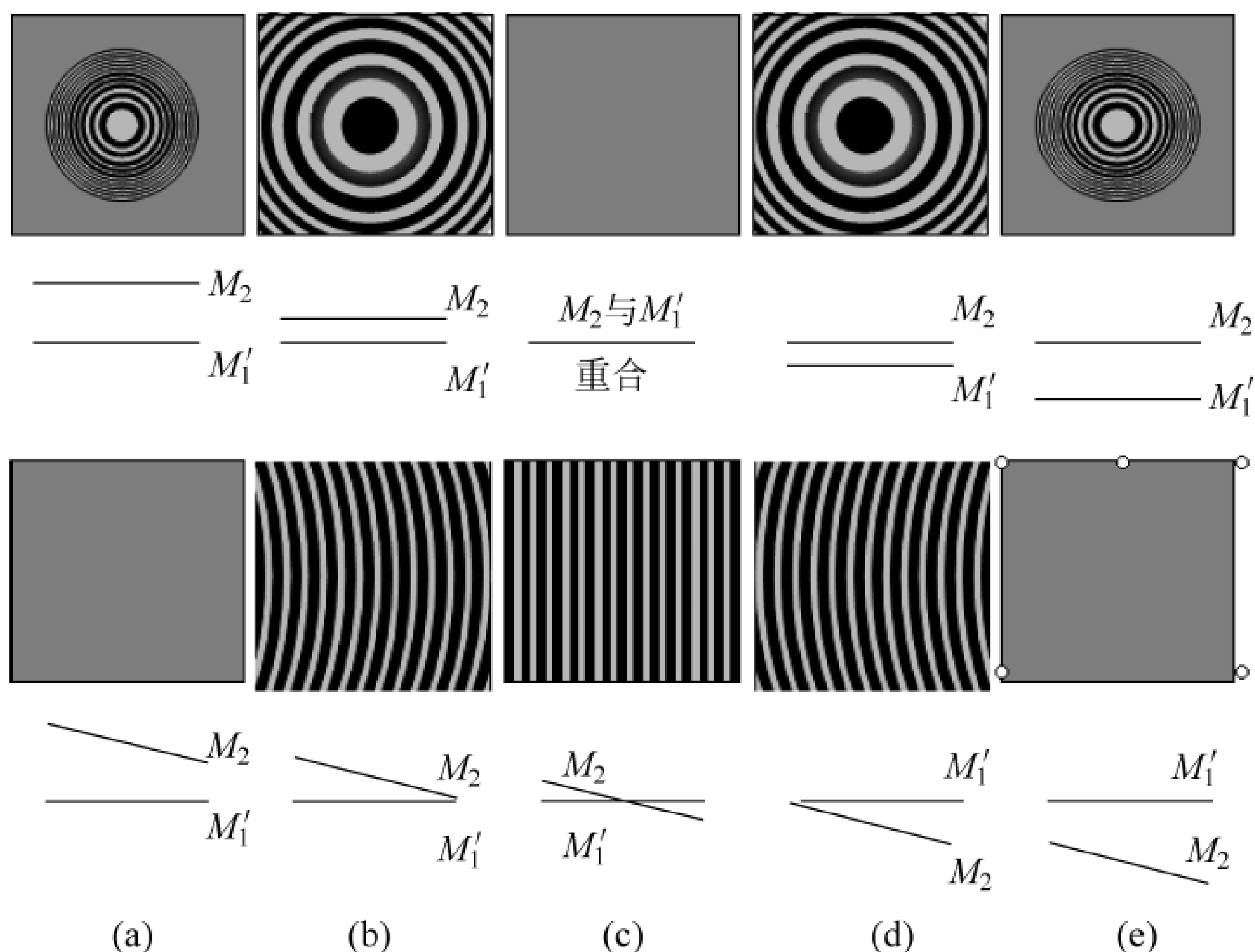



图 8-4-2 迈克耳孙干涉仪的条纹变化

当 $M_1$ 到 $G_1$ 的距离接近于 $M_2$ 到 $G_1$ 的距离（即 $d$ 趋于零）时，只看到均匀亮度，而没有干涉纹，如图 8-4-2(c)所示。$M_1$ 再继续前移，这时 $M_1'$ 在 $M_2$ 的右边，并且距离 $d$ 逐渐增大。和上面 $d$ 减小时相似，这时视场中又出现干涉纹，而且干涉环纹不断从中心涌出，逐渐向外移动，干涉环纹也逐渐变密，如图 8-4-2(d)和(e)。

所以，当 $M_1$ 移动，$d$ 减小时，干涉环纹向内收缩，“陷入”中心，同时，环纹变稀；当 $d$ 增加时，干涉环纹从中心“涌出”，向外扩张，同时环纹变密。无论是“陷入”或“涌出”一个环纹，都表明 $M_1$ 移动了 $\lambda/2$ 的距离。所以，$M_1$ 的位置极微小的变化，都可由干涉条纹的变化进行精密地量度。

在等倾干涉图样中，第 $k+1$ 级亮条纹是由入射角 $i'_{k+1}$ 满足下式的光产生的：

$$2d\cos i'_{k+1} = (k+1)\lambda \tag{8-4-4}$$

同理，第 $k$ 级亮条纹满足

$$2d\cos i'_k = k\lambda \tag{8-4-5}$$

由比较式(8-4-4)和式(8-4-5)可知 $i'_{k+1} > i'_k$，即较高级次 $k+1$ 级的干涉条纹在较低级

次 $k$ 级干涉条纹的内侧，愈向边缘，干涉条纹的级次愈小，这是和牛顿环干涉条纹截然不同的。在同一等倾干涉的干涉图样中，两相邻干涉条纹的间隔可从下述方法求出：

由式(8-4-4)减式(8-4-5)，得：$2d(\cos i'_{k+1}-\cos i'_k)=\lambda$，或写成：$-2\sin\dfrac{i'_{k+1}+i'_k}{2}\sin\dfrac{i'_{k+1}-i'_k}{2}=\dfrac{\lambda}{2d}$，实际上，$i'_k$ 和 $i'_{k+1}$ 的绝对值都很小，所以，上式可近似写成：$-(i'_{k+1}+i'_k)(i'_{k+1}-i'_k)=\dfrac{\lambda}{d}$，或

$$\Delta i'_k=-\frac{\lambda}{2d}\cdot\frac{1}{\bar{i}'_k} \tag{8-4-6}$$

式(8-4-6)中 $\Delta i'_k=i'_{k+1}-i'_k$，$\bar{i}'_k=(i'_{k+1}+i'_k)/2$。此式表明相邻两条纹的角距离$\Delta i'_k$ 正比于波长而反比于 $M'_1$ 与 $M_2$ 之间的间隔，并且随条纹对干涉中心的远离($\bar{i}'_k$ 大)，条纹的间隔越来越小。这一点在表观上与牛顿环的干涉图样相似。图 8-4-2 已清楚地表明了这种情况。

(2) 等厚干涉

当 $M_1$ 与 $M_2$ 不完全垂直，即 $M'_1$ 与 $M_2$ 之间形成一楔形“空气层”时，可得等厚干涉条纹。这时，在空气层上呈现出平行等距离的直条纹。不过，在此情况下，光程差还多少与角度有关，不是严格的等厚条纹，因而条纹有一些弯曲；只有在 $M_2$ 与 $M'_1$ 相距很近或相交时，才呈现很平行的直条纹。由这些干涉图样的变化，可以判断 $M_1$、$M_2$ 的位置。

### 8.4.3 白光干涉及补偿板的作用

(1) 白光干涉

用迈克耳孙干涉仪作精密测量时，需要调节 $M_1$ 处于零光程差的位置，即分光板 $G_1$ 分开的两束光，由 $M_1$ 和 $M_2$ 反射后到达接收器 $E$ 的光程差为零。这时，如果用单色光看等倾干涉环纹，视场将是均匀颜色，没有干涉纹。但是，如果光程差很小，例如 $\Delta=2d=\lambda$，则按照 $2d\cos i=k\lambda$，这时干涉图样中，中央点以外的第一个暗纹的 $i$ 应是，$\cos i=\dfrac{\lambda}{2}/2d=\dfrac{1}{2}$，取 $2d=\lambda$，即得 $i=60°$。按照干涉仪的装置，射到 $M_1$、$M_2$ 上的光线远没有这样大的入射角，所以在视场中也看不到干涉纹，从光强分布上也难以判断和零光程差时有什么区别。因此，在等倾干涉中，只是看到视场是均匀亮度，没有干涉纹，并不能判断一定是处于零光程差的位置。如果用单色光观察等厚干涉条纹，当 $M_2$ 和 $M'_1$ 距离很近或在中点相交时，看到的干涉图样都是均匀的明暗相间直条纹，也无法判断零光程差的位置。

但这时(等厚时)如改用白光，则当 $M_2$ 和 $M'_1$ 相交时，相交处对应的是光程为零的零级条纹，各种波长的光重叠不显色散，而零级条纹左右两旁，将对称分布一级、二级……彩色干涉条纹。

如果 $M_2$ 及 $M'_1$ 不相交时，将看不到这样的干涉图样。所以利用白光看等厚干涉条纹，可以定零光程差的位置。

(2) 补偿板 $G_2$ 的作用

如果没有补偿板 $G_2$，在用白光干涉时，光线在 $G_1$ 板的半反膜分裂成两支，一支透过 $G_1$ 板经 $M_1$ 反射后，再透过 $G_1$ 板到达接收器 $E$；另一支经 $M_2$ 反射后，再经 $G_1$ 反射到达接收器 $E$。则前一支光线，其光程中将包括两次，经过玻璃板 $G_1$ 的光程为 $2nl$($n$ 是 $G_1$ 板玻璃的折射率，$l$ 是光线一次经过 $G_1$ 板的几何路程)；而另一支光线并不包含这样的光程。由

于波长不同，折射率 $n$ 也不同，即光程 $2nl$ 对不同波长的光具有不同的值。因此，对某一波长的光调节到两支光线到达接收器光程差为零，对于其他波长的光，光程差将不为零。这样，不可能调到当白光照射时，对应各种波长的光光程差为零的没有色散的零级条纹。也就难定出 $M_1$、$M_2$ 处于零光程差的位置。如果在射向 $M_2$ 的光程中插入补偿板 $G_2$，则两支光线都两次经历相同的玻璃板。这样，对于某一波长的光，两支光路的光程差为零，对于其他波长的光，光程差也同样为零。所以补偿板只是在白光调节零光程差位置时才是必要的。例如，对于 $\lambda=656\text{nm}$ 和 486.1nm。冕牌玻璃的折射率分别为 1.52 和 1.53。这样，0.5cm 厚的冕牌玻璃板将包含 11619 个 656nm 波长和约 15768 个 486.1nm 波长$\left(\text{由公式 }\lambda=\dfrac{2d_0}{N}\right.$ 算出，$d_0$ 为 $M_1$ 移动的距离，$N$ 为移动的条纹数，后面将有推导$\Big)$。因此很明显，如果从 486.1nm 到 656nm 有一个连续的波长范围，就应该有一块补偿板才能观察到条纹图样。

### 8.4.4　使用迈克耳孙干涉仪的注意事项

**1. 光源机械稳定性对条纹的影响**

作为迈克耳孙干涉仪的光源，不管是用激光，还是用钠光灯作为光源，一般有一定的振动，在振动轻微的情况下对观察干涉现象影响不大。如果振动较大，或其中电源的工作不稳定，发出较大的“嗡嗡”声，就会使干涉条纹产生抖动，或使视场产生闪烁，忽明忽暗。对于振动，可用一只手按住光源外壳，减少振动幅度。对于工作电源不稳定“嗡嗡”声过大，必须修理或更换电源。如果用激光器作为光源，须用扩束镜，使视场范围增大，观察到更多的条纹；如果用钠光灯作光源，则可除去原来的灯罩，改用白纸筒作罩，套在钠光灯管上，并在纸筒上画一黑色的十字叉丝作为物的标志，比用毛玻璃的效果更好，操作更方便。

对于双线钠光，干涉条纹的清晰度取决于条纹的粗细和非相干叠加两个方面；对于激光，单色性好，不存在非相干叠加问题。

**2. 影响干涉条纹粗细的因素**

首先，迈克耳孙干涉仪的平面镜 $M_2$ 与 $M_1$ 的虚像 $M_1'$ 之间的距离 $d$，对干涉条纹的粗细清晰度有直接的影响。由式(8-4-6)$\Delta i_k'=-\dfrac{\lambda}{2d}\cdot\dfrac{1}{\bar{i}_k'}$可知影响干涉条纹粗细的因素：一是第 $k$ 级的入射角 $i_k$ 的大小，二是 $M_2$、$M_1'$ 之间的距离 $d$ 的大小。对于视场中心 $i_k$ 一定的范围内，$d$ 越小，即 $M_2$、$M_1'$ 越靠近，条纹的宽度就越大，就越便于观察。反之，$d$ 越大，条纹宽度就越细，不便观察，难以区分。所以，不管是激光还是钠光，当视场中心的干涉条纹很细、难以区分时，就要减小 $d$ 值，才能使条纹变粗。

其次，非相干叠加对干涉条纹清晰度有影响。由于不同波长的光要产生各自独立的干涉条纹，形成同一空间不同波长的干涉光强的非相干叠加。当波长为 $\lambda_1$ 的亮纹与 $\lambda_2$ 的亮纹叠加，则视见度最佳；当 $M_1$ 的位置改变时，即 $d$ 改变，$\lambda_1$ 的亮纹与 $\lambda_2$ 的亮纹就会错开，当 $\lambda_1$ 的亮纹与 $\lambda_2$ 的暗纹叠加，则视见度最差，并发生周期性的变化，所以视见度会发生周期性的变化，由最佳到最差，由最差到最佳。在视见度最差时，干涉条纹很模糊，甚至分不出干涉条纹。

（1）当 $d$ 的值同时分别为 $\lambda_1$、$\lambda_2$ 半波长整数倍，即

$$d=m\lambda_1/2=n\lambda_2/2\tag{8-4-7}$$

则视见度最佳。

(2) 当 $d$ 为 $\lambda_1$ 的半波长的整数倍,而为 $\lambda_2$ 的半波长整数加 0.5 倍,即

$$d=m\lambda_1/2,\quad 同时\ d=(n+0.5)\lambda_2/2 \tag{8-4-8}$$

则此时 $\lambda_1$ 的亮纹与 $\lambda_2$ 暗纹叠加,$\lambda_1$ 的暗纹与 $\lambda_2$ 的亮纹叠加,视见度最差。

(3) 当 $d$ 接近零时,可避免第二种情况的发生,即视见度较佳。由此可见:当 $d\approx 0$ 时,不但条纹宽度最粗,而且条纹视见度最佳。

(4) 当 $d\approx m\lambda_1/2\approx n\lambda_2/2$ 时,条纹视见度较佳,但条纹的宽度不一定足够。所以在迈克耳孙干涉仪的调节过程中,首先最重要的一步就是用目测,使 $M_1G_1$ 与 $G_1M_2$ 大致相等,但不会精确相等。从 $G_1$ 镀膜面算起。

**3. 测量 $d$ 和数条纹过程中应注意的几个问题**

准确测量 $d$ 和数对应条纹的变化数目,是实验的关键。采用下列方法,可以提高测量 $d$ 和数条纹数目的准确度。

(1) 如果是数干涉条纹“涌出”的条数,必须转动微动手轮到有条纹“涌出”时,才能记下 $M_1$ 的初位置,然后将微动手轮朝着相同的方向转动,中间不能反转,直至有 100 条干涉条纹从中心“涌出”时,记下 $M_1$ 的终位置,如果是数“陷入”的条纹数,须先返转粗动手轮,后转微动手轮,至均有条纹“陷入”,方可记下 $M_1$ 的初位置;朝相同方向,数 100 条“陷入”的条纹后记下 $M_1$ 的终位置。

(2) 在数条纹“涌出”或“陷入”的过程中,如果头部发生上下或左右移动,则数的条纹个数与实际因 $M_1$ 改变而“陷入”或“涌出”的条纹的个数不相等,因此而带来粗大误差。为了克服这一原因带来的粗大误差,用左手撑住脸部,可控制头部眼睛上下左右的移动,也可减少疲劳程度。迈克耳孙干涉仪是比较精密的仪器,必须理解它的原理,掌握 $M_1$、$M_2$、$G_1$ 三者与波长以及干涉现象的关系,耐心细致地操作,才能快速调出干涉条纹,准确测出相应的物理量。

### 8.4.5 迈克耳孙干涉仪的应用

迈克耳孙干涉仪的最大特点是相干两束光完全分开,在光路中可插入待研究的介质,可以按需要调节光程差,其精确量度可达几十分之一波长,甚至百分之一波长。在精密测量方面,有广泛的应用。

迈克耳孙设计的干涉仪,最先是用来研究“以太”性质。他和莫雷合作,企图测量地球相对“以太”运动的速度。这一实验称做迈克耳孙-莫雷实验,是近代物理学发展中的关键性实验之一。

在用迈克耳孙干涉仪观察等倾干涉条纹时,对中央条纹,$M_1$ 移动 $d_0$ 时有 $N$ 条条纹收缩到中心,则由 $\begin{cases}2d=k\lambda\\2(d-d_0)=(k-N)\lambda\end{cases}$,可得

$$\lambda=\frac{2d_0}{N} \tag{8-4-9}$$

迈克耳孙干涉仪还可以测量两种相近波长的波长差。假定有一盏钠光灯,它主要发出两种相近的波长 589.0nm 和 589.6nm。首先将干涉仪调节在零光程差位置(利用白光的等厚干涉图样寻找零光程差位置,这时只能看到 $d=0$ 附近的几个彩色条纹),在接近 $d=0$ 的情况下,两组条纹图样重叠在一起。如果把平面镜 $M_1$ 移离(或移向)板 $G_1$ 一段距离 $d$,则对应于波长 $\lambda_1$ 的极大值一般不与 $\lambda_2$ 的极大值有相同的位置。事实上,如果 $d$ 满足 $\frac{2d}{\lambda_1}-\frac{2d}{\lambda_2}=\frac{1}{2}$,并且 $2d\cos i'=k\lambda_1$,则 $2d\cos i'=\left(k+\frac{1}{2}\right)\lambda_2$。这样,$\lambda_1$ 的极大值就落到 $\lambda_2$ 的极小

值上，$\lambda_1$ 的极小值落到 $\lambda_2$ 的极大值位置上。所以条纹系将消失。容易看出，若$\frac{2d}{\lambda_1}-\frac{2d}{\lambda_2}=1$，则干涉图样再次出现。一般来说，若$\frac{2d}{\lambda_1}-\frac{2d}{\lambda_2}$等于 1/2，3/2，5/2，…，则干涉图样消失；若$\frac{2d}{\lambda_1}-\frac{2d}{\lambda_2}$等于 1，2，3，…，则条纹图样出现。若 $\lambda_2$ 用 $\lambda_1+\Delta\lambda$ 表示，则可以求出两种靠近波长的波长差。

迈克耳孙干涉仪活动镜 $M_1$ 移动的范围可达几十厘米。按 $2d\cos i'=k\lambda$，$d$ 大则 $k$ 也大(例如 $k=10^8$)。由于干涉级次 $k$ 很大，如果一束光包含两种波长差很小(例如 0.1nm)的光射到迈克耳孙干涉仪，在干涉图样上便可以看出两组干涉环纹，从而测出各自的波长。所以，迈克耳孙干涉仪可用于研究光谱线的精细结构。

迈克耳孙干涉仪可用以精密测定介质的折射率，或折射率的微小改变。在迈克耳孙干涉仪的一支光路中，插入与空气折射率不同的介质，若介质折射率为 $n$(待测)，长为 $l$，则两支光路的光程差改变了 $2(n-1)l$。由于干涉条纹变化的数值和光波波长，可以计算此光程差的改变，从而测出介质的折射率。

按照迈克耳孙干涉仪的基本原理，后人加以发展和改进，制成各种形式的干涉仪，其中有著名的泰曼(Tyman)干涉仪和傅里叶分光计，应用到光学研究的各个方面。

## 8.5　多光束干涉与自由光谱程

在实际应用中，干涉图样最好是十分狭窄、边缘清晰并且十分明亮的条纹，此外还要求亮条纹能被比较宽阔且相当黑暗的区域隔开，因此，需要采用相位差相同的多光束干涉系统才能满足以上要求。从数学角度上来讲双光束干涉方程为 $I=4A_1^2\cos^2\frac{\Delta\varphi}{2}$。当 $\Delta\varphi=2k\pi$ 时 $I=4A_1^2$，$\Delta\varphi=(2k+1)\pi$ 时，$I=0$。当 $\Delta\varphi$ 处于 $2k\pi$ 和$(2k+1)\pi$ 之间时，$0<I<4A_1^2$，由于 $I$ 随 $\Delta\varphi$ 的变化比较缓慢，所以干涉条纹本身比较宽，不够细锐。为使条纹细锐，人们设计了法布里-珀罗干涉仪，实现多光束干涉，它能形成非常明亮、细锐清晰的同心圆形的等倾干涉条纹。

### 8.5.1　法布里-珀罗干涉仪的组成结构和工作原理

(1) 组成结构

法布里-珀罗干涉仪主要由平行放置的两块平面板组成，入射光进入两镜之间，经过多次反射，在第二个镜的透射方向产生多光束。由于这些多光束实际是同一束光分解来的，所以是相干的。如图 8-5-1 所示，在两个板相向的表面 G 和 G′上镀有薄银膜或其他反射率较高的薄膜，要求镀膜的表面与标准样板之间的偏差不超过 1/50～1/20 个波长，为了消除两平板相背的平面上的反射光的干涉与所研究干涉的重叠，每块板都不是平行面板，板的两面成一很小的夹角，如图 8-5-2 所示。若两平行的镀银平面的间隔用由某些热膨胀系数很小的材料做成的小环固定起来，则称该仪器为法布里-珀罗标准具；若两平行的镀银平面的间隔可以改变，则称为法布里-珀罗干涉仪。

(2) 工作原理

如图 8-5-1 所示，将扩展面光源 $S$ 放在透镜 $L_1$ 的焦平面上，使许多方向不同的平行光束入射到干涉仪上，自光源 $S$ 上任一点发出的光，射在两平板后，则在两镀银的平面间进行

多次反射,结果构成多个平行的透射光1,2,3,…和多个平行的反射光1′,2′,3′,…在这两组光中,相邻光的相位差和光程差都相等,设入射振幅为 $a$,则透射光的振幅分别为 $att'$,$att'r^2$,$att'r^4$,…而反射光的振幅为 $ar$,$artt'$,$ar^3tt'$,$ar^5tt'$,…(见图8-5-2),其中 $r^2$ 为镀银面的反射系数,因为 $r^2+tt'=1$,当反射系数 $r^2$ 较大且与1相差不多时,构成的透射光接近于等振幅的多光束,这样透射的多光束,就可作为获得细锐的干涉条纹的 $\lambda$ 射光。

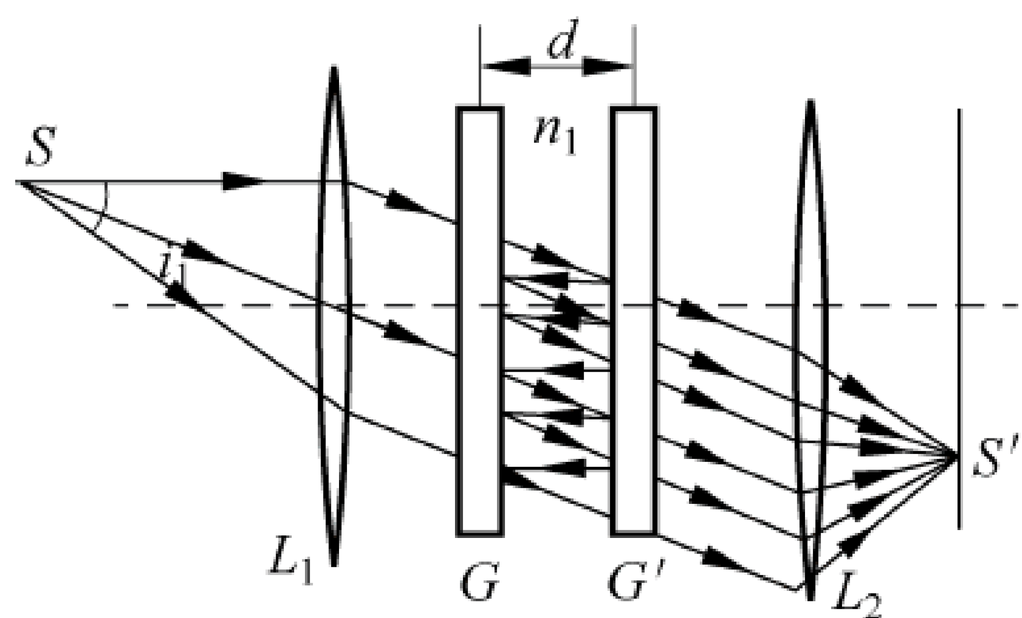



图8-5-1 法布里-珀罗干涉仪的组成结构

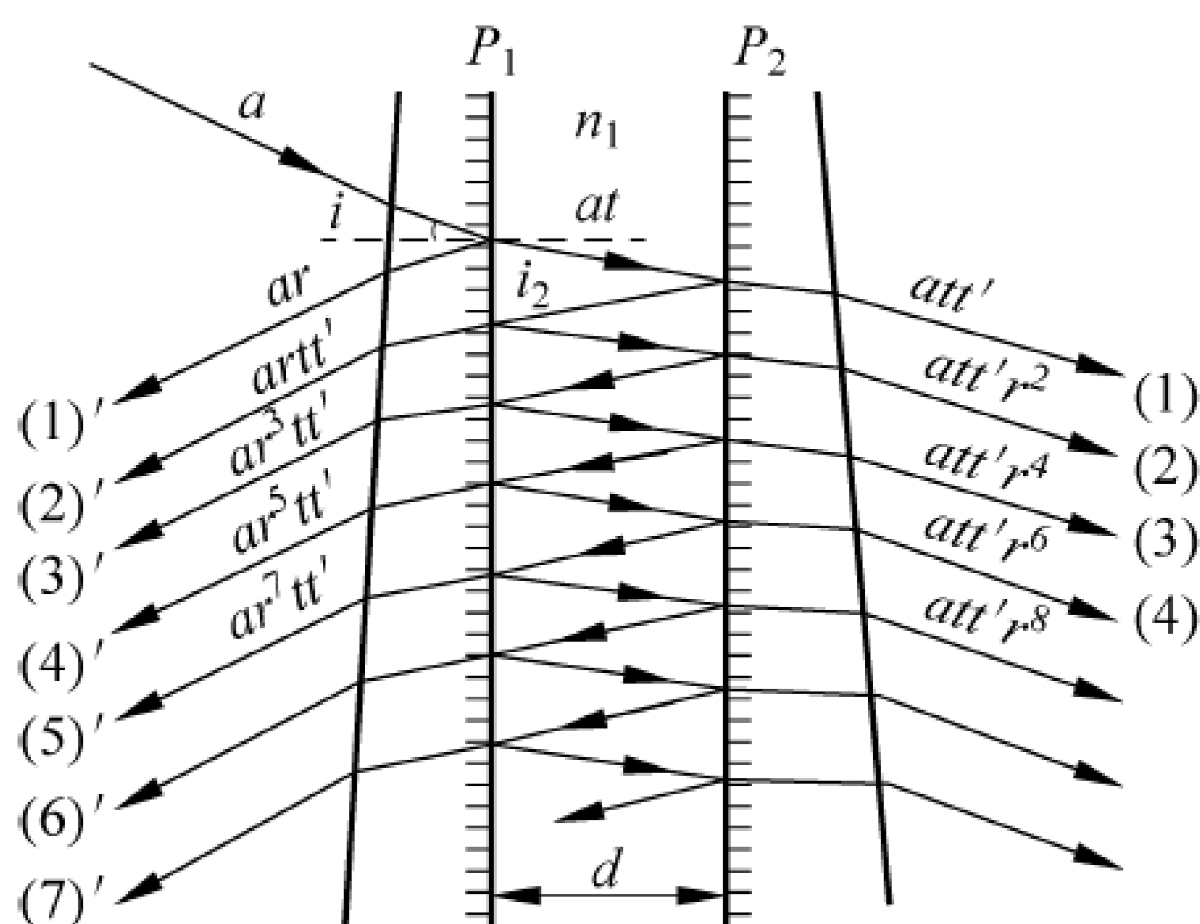



图8-5-2 法布里-珀罗干涉仪的工作原理

如图8-5-1所示,在透射的光束中,相邻光束的光程差为 $\delta=2n_1d\cos i_2$,由此引起的相位差为

$$\varphi=\frac{2\pi}{\lambda}\delta=\frac{4\pi}{\lambda}n_1d\cos i_2 \tag{8-5-1}$$

由于一般情况下,法布里-珀罗干涉仪是在空气中使用,空气的折射率为 $n_1=1$,所以,上两式可写为 $\delta=2d\cos i_2$ 和 $\varphi=\frac{2\pi}{\lambda}\delta=\frac{4\pi}{\lambda}d\cos i_2$。

当此光程差为波长的整数倍时,则得到最大光强度,称为干涉主最大,由于两镀银平面是平行的且所用的光源为扩展面光源,所以产生等倾干涉,在有相同入射角的光线所产生的光束中,相邻光线的相位差相同,而入射角相同的光线在垂直于观察方向的平面上的轨迹,是一组同心圆,因而在透镜 $L_2$ 焦平面上得到的是一组同心环形条纹的图样。

### 8.5.2 关于法布里-珀罗干涉图样的特点的分析

透过法布里-珀罗干涉仪的光在透镜 $L_2$ 焦平面上形成细锐的干涉亮条纹,这个特点使它成为研究光谱线超细结构的有效手段。为了进一步了解法布里-珀罗干涉图样的特点,首先分析干涉亮条纹的光强分布和表征光强分布的特性的两个参量。

(1) 透射光在透镜 $L_2$ 焦平面上的光强的分布

考虑到传播方向垂直于彼此平行的平面镜的平面波,两个平面镜的距离为 $d$,假定它们具有相同的反射率和相同的透射率,设入射平面波为 $E_0=E_e(z,t)=E_{e0}\exp(ikz-i\omega t)$,经过一次反射的反射波为 $E_r$,经过一次透射的透射波为 $E_t$,为简单计,暂不考虑由反射和透射引起的附加相位移,因为它们不会改变仪器的工作原理,这样,振幅反射率 $r$ 和振幅透射率 $t$ 分别定义为:$r=E_r/E_e$ 和 $t=E_t/E_e$,它们是0~1的实数。假设反射镜是无吸收的,根据能量守恒,有 $r^2+t^2=1$。

由以上假设不难得到入射波 $E_e$ 透过第一个反射镜后的电场振幅:$E_t=tE_e(0,t)=tE_{e0}$ 和光波第一次透过第二个反射镜后的电场振幅:

$$E_1 = tE_t\exp(ikd) = E_{e0}t^2\exp(ikd) \tag{8-5-2}$$

因为反射镜之间的光程为 $d$，所以式(8-5-2)中引入了相位因子 $\exp(ikd)$，光波 $E_t\exp(ikd)$的一部分入射在第二个反射镜上并被反射：$E_{1r} = \mathrm{tr}E_{e0}\exp(ikd)$。它返回到第一个反射镜，再经过第一个反射镜的反射并再次透过第二个反射镜后，其振幅为

$$E_2 = E_{e0}t^2r^2\exp(ik3d) \tag{8-5-3}$$

该输出波经受了两次反射($r^2$)和两次透射($t^2$)，因为光程为 $3d$，则增加了相位移 $\exp(ik3d)$。由式(8-5-2)、式(8-5-3)可以看出透射波振幅之间有一个关系，即每往返一次后，波振幅便乘以一个 $r^2\exp(ik2d)$，即 $E_n = E_{n-1}r^2\exp(ik2d)$，$n=2,3,4,\cdots$或者 $E_n = E_{e0}t^2\exp(ikL)r^{2(n-1)}\exp[ik2(n-1)d]$。

这样，在第二个反射镜后面的多个平面波 $E_1, E_2, E_3, \cdots$的叠加，便得到了输出波振幅：

$$\begin{aligned}
E_a &= \sum_{n=1}^{\infty}E_n = E_{e0}t^2\exp(ikd)\sum_{n=1}^{\infty}r^{2(n-1)}\exp[ik2(n-1)d] \\
&= E_{e0}t^2\exp(ikd)\sum_{n=1}^{\infty}r^{2n}\exp(ik2nd) \\
&= E_{e0}t^2\exp(ikd)\sum_{n=1}^{\infty}[r^2\exp(ik2d)]^n
\end{aligned} \tag{8-5-4}$$

这个等比数列的和可以表示为 $\sum\limits_{n=1}^{\infty}q^n = \dfrac{1}{1-q}$(对于$|q|<1$)。这里，$q = r^2\exp(ik2d)$。所以，干涉仪的输出振幅为

$$E_a = E_{e0}t^2\exp(ikd)\frac{1}{1-r^2\exp(ik2d)} \tag{8-5-5}$$

由于能看到的和测量到的只能是光强，因此必须给出输出强度的表达式，为了简化，引进相位差 $\varphi$，它表示在反射镜之间往返一次的相位差：$\varphi = k2d$。则输出强度为

$$\begin{aligned}
I_t &= E_aE_a^* = E_{e0}^2t^4\frac{1}{(1-r^2\mathrm{e}^{\mathrm{i}\varphi})(1-r^2\mathrm{e}^{-\mathrm{i}\varphi})} \\
&= I_et^4\frac{1}{1+r^4-r^2\mathrm{e}^{-\mathrm{i}\varphi}-r^2\mathrm{e}^{\mathrm{i}\varphi}} \\
&= I_et^4\frac{1}{1+r^4-2r^2\cos\varphi}
\end{aligned} \tag{8-5-6}$$

这里 $I_e = E_{e0}^2$ 是干涉仪的入射光强度，也可以写为

$$\begin{aligned}
I_t &= I_et^4\frac{1}{1+r^4-2r^2(1-2\sin^2(\varphi/2))} \\
&= I_et^4\frac{1}{(1-r^2)^2+4r^2\sin^2(\varphi/2)}
\end{aligned} \tag{8-5-7}$$

由 $r^2+t^2=1$，得

$$I_t = \frac{I_e}{1+F\sin^2(\varphi/2)} \tag{8-5-8}$$

而 $F = \left(\dfrac{2r}{1-r^2}\right)^2$ 为精细度，反映干涉条纹的细锐程度。而 $T_1 = \dfrac{I_t}{I_e} = \dfrac{1}{1+F\sin^2kd}$，为

该干涉仪的强度透射率,称之为爱里函数,其图像如图8-5-3所示。由图可以看到,当$r^2(R)$接近1时,透射函数曲线尖锐,此时透射光干涉图样由几乎全黑的背景下一组很细的亮条纹构成,随着反射率的增大,透射光暗条纹的强度降低,亮条纹的宽度变窄,亮条纹越加细锐,因此,条纹的锐度和可见度增大。

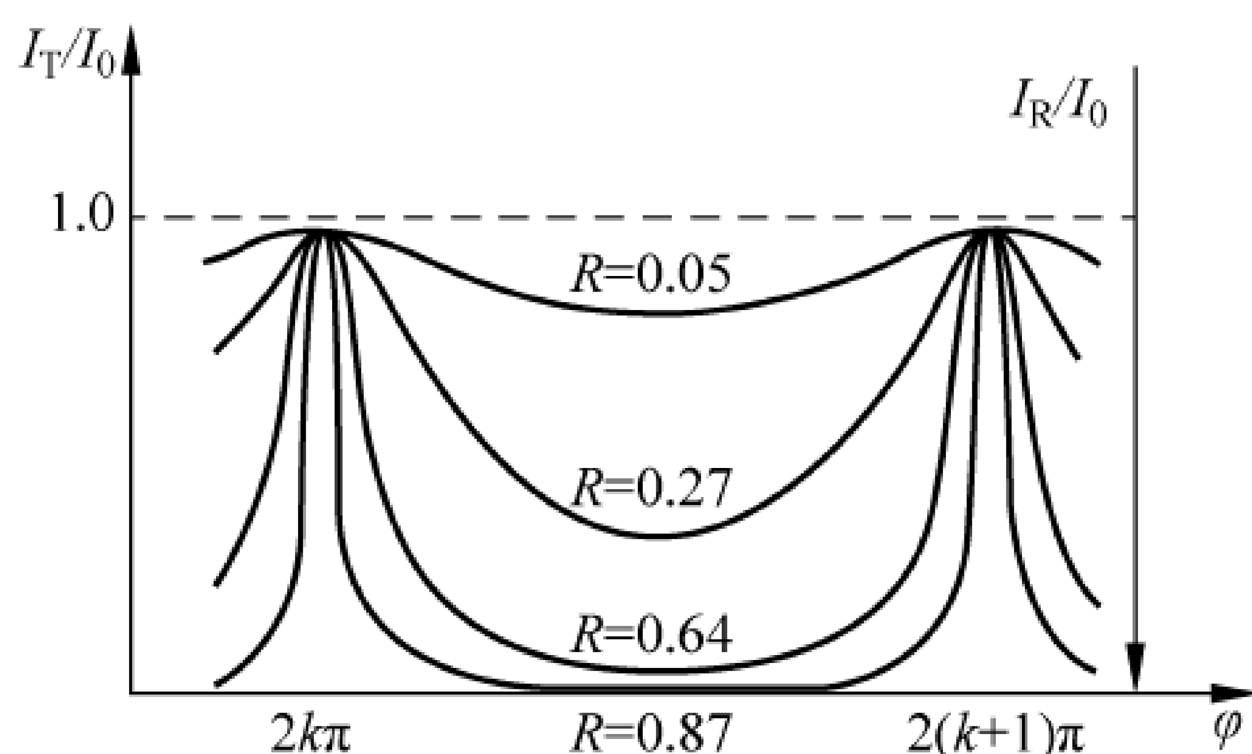


图8-5-3 多光束干涉光强分布图

$T_1$是爱里函数是$kd$的周期函数,如果$d$固定不变,只改变$k=2\pi/\lambda$或波长$\lambda$,则会周期性地观察到相同的强度透射率,当正弦函数$\sin kd=0$时,艾里函数$T_1=1$,这对应于某个波长或频率的周期性的完全透射,因此,干涉仪可以对波长或频率进行选择,因而,法布里-珀罗标准具可以用作光学滤波器。

(2)自由光谱范围

自由光谱范围是指两个透射亮条纹最大值(主最大)之间的波长差。

设波长为$\lambda_1$和$\lambda_2(\lambda_1>\lambda_2)$的两光以相同的方向射到法布里-珀罗标准具上,它们各自产生一组同心圆环状的干涉亮条纹(主最大),对同一干涉级($k$),$\lambda_2$的干涉圆环的直径较$\lambda_1$的小些,当满足$2d\cos i'=k\lambda_1=k\lambda_2$时,$\lambda_1$的第$k$级亮圆环与$\lambda_2$的第($k-1$)级亮圆环重叠,因而得

$$\Delta\lambda=\lambda_1-\lambda_2=\frac{\lambda_2}{k} \tag{8-5-9}$$

由于在法布里-珀罗干涉标准具中,大多数情况下,$\cos i_2\approx 1$,所以式(8-5-9)中的$k$值应为$k\approx\dfrac{2d}{\lambda}$。以此$k$值代入式(8-5-9)得$\Delta\lambda=\dfrac{\lambda_1\lambda_2}{2d}$。实际上,可以认为$\lambda_1\lambda_2=\lambda_1^2$,还可以省略$\lambda$的下标,故有

$$\Delta\lambda=\frac{\lambda_1^2}{2d} \tag{8-5-10}$$

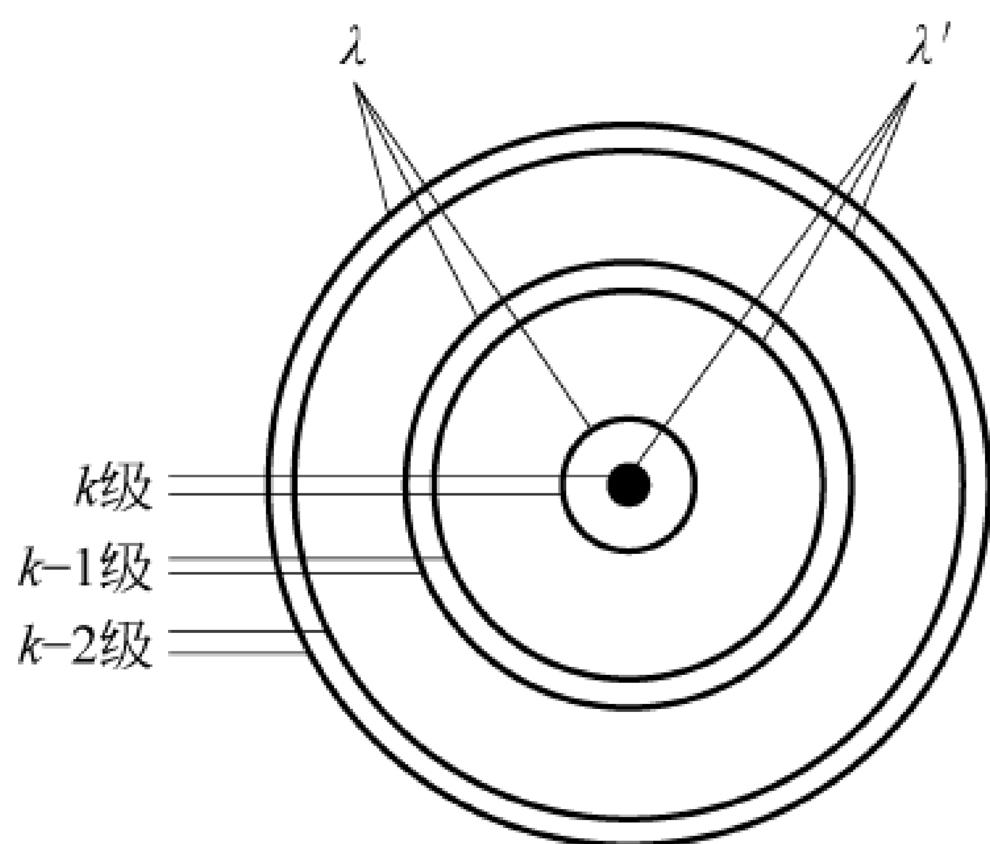


图8-5-4 标准具的自由光谱范围

此$\Delta\lambda$值是某一波长光的干涉圆环和另一波长光的干涉圆环重合时的波长差,亦即在给定$d$的标准具中,若入射光的波长在$\lambda_1$到$\lambda_1+\Delta\lambda$的波长范围以内,则所产生的干涉圆环不重叠,则称此$\Delta\lambda$为标准具的自由光谱范围。而当干涉环不越级重叠时,才能达到光源中光谱间隔的最大限度。因此,我们可以根据式$\Delta\lambda=\dfrac{\lambda_1^2}{2d}$,通过波长$\lambda_1$求出法布里-珀罗干涉具的自由光谱范围(图8-5-4)。

### 8.5.3 分辨本领

前面已分析过,法布里-珀罗干涉标准具的镀银面的反射系数越大,由透射光所得的干涉亮圆环亦越细锐,因而刚能被分辨或刚能鉴别的两相邻亮环的几何间隔越小,亦即刚能被分辨的相应的两相邻波长的波长差$\Delta\lambda$越小,简言之,标准具的反射系数$R(r^2)$越大,则刚能被分辨的两亮圆环所对应的两光的波长差$\Delta\lambda$越小,$\lambda/\Delta\lambda$为光谱分辨本领,所以法布里-

珀罗标准具的光谱分辨本领与镀银面的反射系数密切相关,反射系数越大,分辨本领亦越大。且

$$R_0=\frac{\lambda}{\Delta\lambda}=kN_e \tag{8-5-11}$$

式中 $k$ 是级数,因 $2d=k\lambda$,间隔 $d$ 越大,则级数越高,分辨本领也越大,$N_e$ 为有效光束的数目,它随反射面的反射系数 $R$ 而增加,为获得高分辨率本领,$R$ 须在 90%以上。例如,对于 $\lambda=564.1\text{nm}$,$d=5\text{mm}$ 的标准具,当反射系数 $R=0.9$ 时,可得光谱分辨本领 $\lambda/\Delta\lambda\approx5.5\times10^5$,这就是对 564.1nm 的光能分辨的波长为 0.1nm,可见,法布里-珀罗标准具是一种分辨本领很高的光谱仪器,也正因如此,它才能用来研究单个光谱线的超精细结构,测单个光谱线的超精细结构就是对粗看来是一条光谱线的光测出它的各光谱成分的波长和强度。

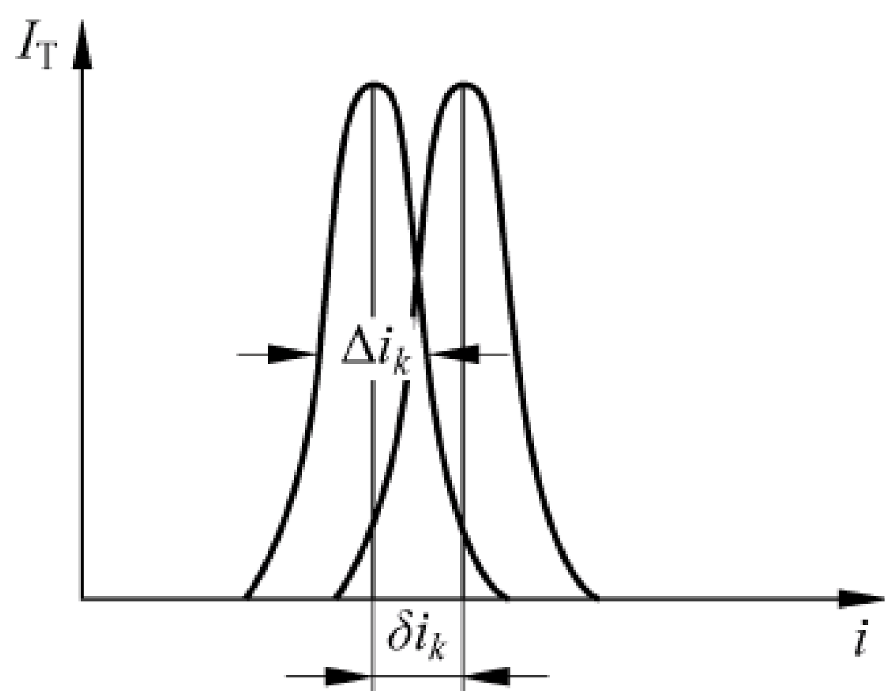



图 8-5-5 标准具的分辨极限

## 8.5.4 法布里-珀罗干涉仪的应用

由于法布里-珀罗干涉仪利用多光束干涉原理能够产生半宽度非常小的细锐条纹,具有极高的色分辨本领,能够精确测定波长产生的干涉条纹非常细锐,因此它一直是长度计量和研究光谱超精细结构的有效工具,并在用波长作长度基准的研究中得到应用,此外它还是激光共振腔的基本结构,其干涉理论也是研究干涉滤片的基础,在光学中一直起着重要的作用。

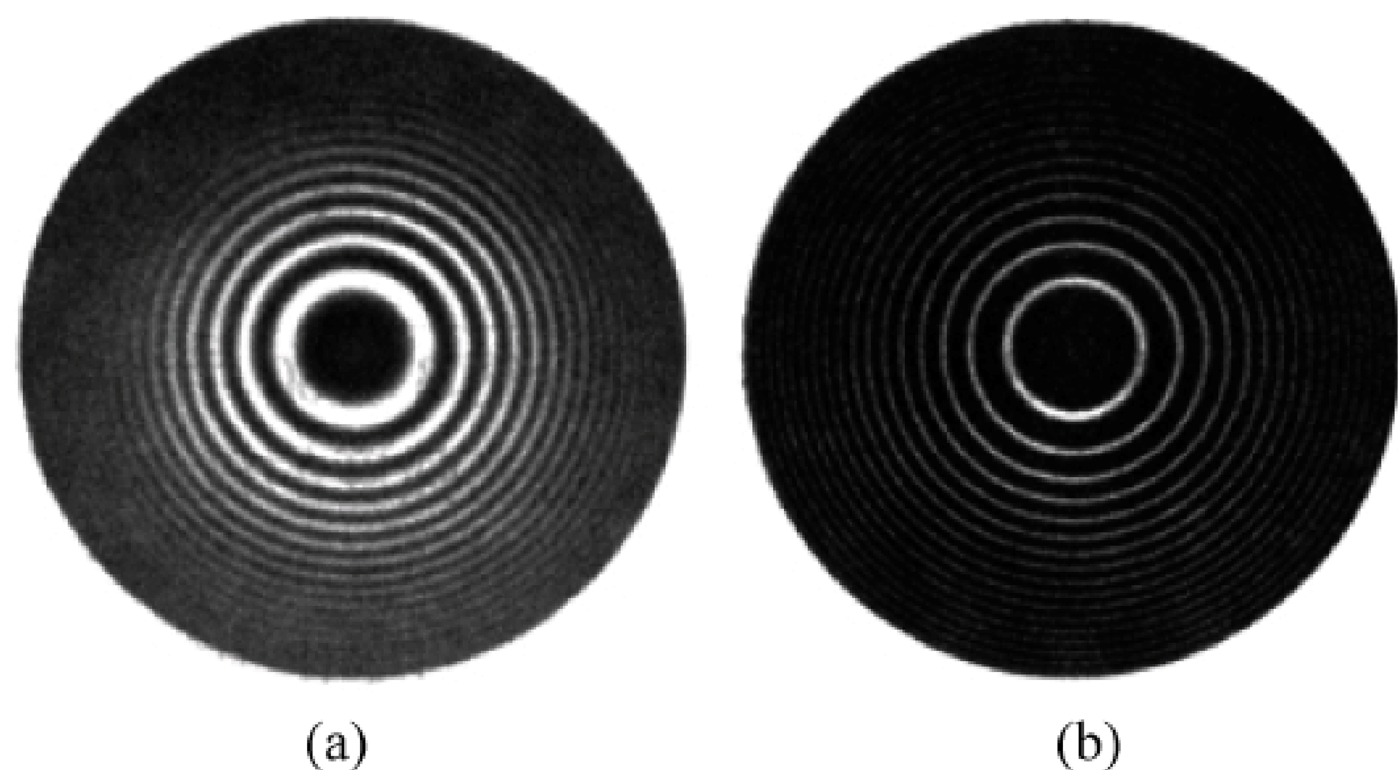



图 8-5-6 迈克耳孙干涉仪(a)和法布里-珀罗干涉仪(b)条纹的比较

这里主要介绍法布里-珀罗腔的应用。

法布里-珀罗腔最早用于光谱精细结构的研究。如果以单色扩展光入射,则 $k$ 固定,但有各种可能的倾角 $i$,只有特定的方向上出现干涉极强。在 $l$、$n$ 不变的情况下,可以推出第 $k$ 级亮纹角宽度 $\Delta i_k$ 为

$$\Delta i_k=\frac{\lambda}{2\pi nl\sin i_k}\frac{1-R}{\sqrt{R}} \tag{8-5-12}$$

从式(8-5-12)可以看出反射率 $R$ 值越高,腔长 $l$ 越大,条纹越细锐。采用按法布里-珀罗干涉仪原理制成的长腔结构,可以得到非常细锐的条纹。如取,$R\approx0.9$,$l=5\text{cm}$,$\lambda\approx0.6\mu\text{m}$,$n\sin i_k\approx1/2$,则角宽度约为 $0.001'$ 量级。如此细锐的条纹使人们有可能更精密地测定它们的精确位置,比较各相近谱线波长,以及开展用波长度量长度的工作。

法布里-珀罗腔的另一种应用是作为激光器的谐振腔。如果以非单色平行光入射,则 $i$ 固定(比如,$i=0$)。由于多光束干涉,使得在很宽的光谱范围内只在某些特定的波长附近出现极大,且相邻极强的频率间隔相等,间隔与腔长成反比。每条谱线称为一个纵模,第 $k$ 级纵模的谱线宽度为

$$\Delta\lambda_k=\frac{\lambda^2}{2\pi nl\sin i}\frac{1-R}{\sqrt{R}}=\frac{\lambda}{\pi k}\frac{1-R}{\sqrt{R}} \tag{8-5-13}$$

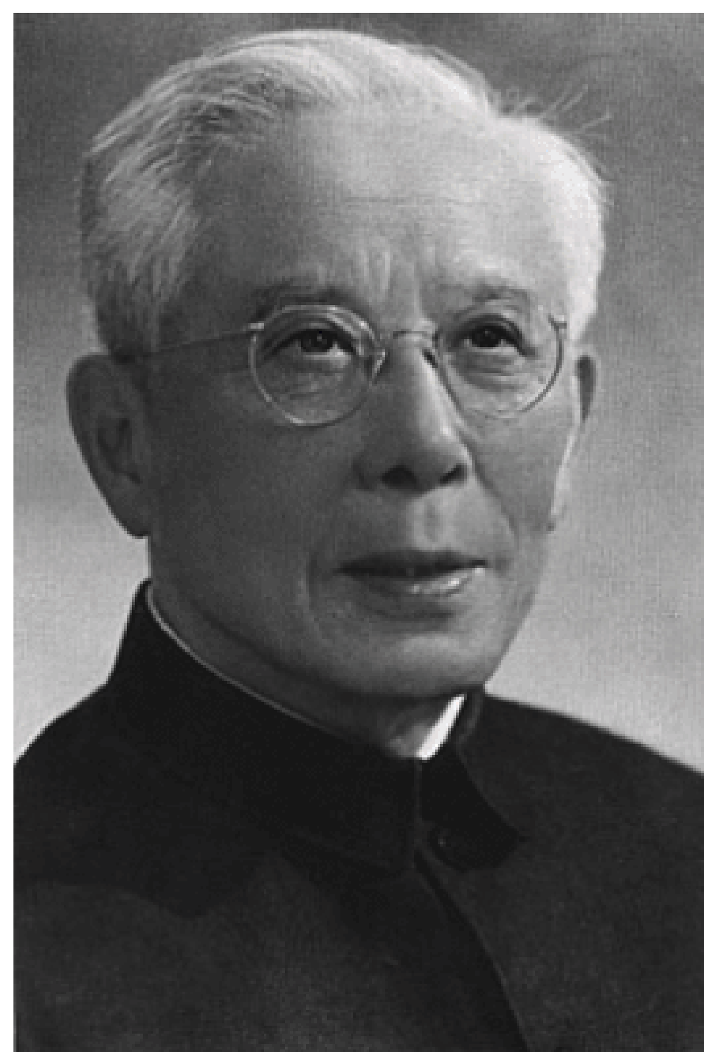
从上式可以看出,反射率越高,腔越长,则谱线宽度越窄。法布里-珀罗腔的这一特性起着挑选波长,压缩线宽,从而提高单色性的作用,在激光技术中得到广泛应用。如利用 $L=20\text{cm}$,$R=98\%$ 的法布里-珀罗腔可以将 He-Ne 激光($\lambda=632.8\text{nm}$)单模线宽压缩至 $10^{-6}\text{nm}$ 量级,加上激光器本身是一个振荡源,其单模线宽远比 $\Delta\lambda_k=\frac{\lambda^2}{2\pi nl\sin i}\frac{1-R}{\sqrt{R}}=\frac{\lambda}{\pi k}\frac{1-R}{\sqrt{R}}$ 式所给出的数值小,例如对于功率为 1mW,腔长为 100cm,反射率为 98%,反转度为 100 的 He-Ne 激光器,其单膜线宽理论上可达 $\Delta\lambda=10^{-13}\text{nm}$ 量级。

总之,法布里-珀罗干涉仪是一种能够实现多光束干涉的重要实验装置,它是一种分辨率非常高的光谱仪器,可以精确地比较各光谱线的波长,这些特殊的价值使得它在近代现代物理以及光学各领域具有广泛的应用。

# 学习资源

## 一、中国近现代著名科学家(物理学)简介

周培源(1902—1993 年),江苏省宜兴人,中共党员。力学家、理论物理学家、教育家和社会活动家,中国近代力学事业的奠基人之一。培养了几代知名的力学家和物理学家。



周培源出生于江苏省宜兴县的一个书香门第,其父是清朝秀才。中学时期在上海圣约翰大学附属中学学习。1919 年考入清华学校(今清华大学前身)中等科,1924 年清华学校高等科毕业。1924 年赴美国,在芝加哥大学学习。1926 年获学士、硕士学位。1927 年到美国加利福尼亚理工学院攻读研究生,做相对论方面的研究。1928 年获博士学位。1928 年秋,赴德国莱比锡大学,在海森伯(Heisenberg)教授指导下从事科学研究;1929 年赴瑞士苏黎世联邦工业大学,在泡利(Pauli)教授指导下从事理论物理研究。1929 年回国后,任清华大学物理系教授。

周培源代表作品:《三等分角法二则》《爱因斯坦引力论中旋转对称物体的引力场》等。他将自己精力的大部分献给了力学与理论物理中两个十分困难的领域:湍流理论和广义相对论。他先后发表了数十篇论文,在这两个领域中都取得了世人瞩目的成就。周培源是一位著名的社会活动家,他曾兼任过许多重要的社会职务。由于周培源在科学研究、教学和社会活动中取得的成就,他受到国内外科学界和教育界的尊敬。1980 年,美国普林斯顿高等学术研究院授予他名誉法学博士学位。1980 年和 1985 年,他曾两次获美国加利福尼亚理工学院"具有卓越贡献的校友"奖。在教育和科学研究中,一贯重视基础理论,同时关怀和支持新技术的研究。在组织领导中国的学术界活动、推进国内外交流合作方面做出了重要贡献。

## 二、诺贝尔物理学奖简介

**1961 年诺贝尔物理学奖**。用直线加速器从事高能电子散射研究并发现核子。获奖者:R.霍夫斯塔特(美国)。从事 $\gamma$ 射线的共振吸收现象研究并发现了穆斯保尔效应。获奖者:R. L.穆斯保尔(德国)。评价:有时候科学成果还真是靠钱烧出来的,理论上直线加速器虽然可以使粒子无限接近光速,不过那成本是相当不菲。霍夫斯塔特由于在斯坦福所以就有

使用这种仪器的机会。当然能拿奖还是他有真才实学(否则斯坦福那么多人怎么不得奖呢),他利用高能粒子测量出了原子核内部的质子和中子存在某种结构分布不是人们想象的为物质的基本结构(原子核带正电,当距离近到一定程度时,由于同性相斥会把质子弹开,但能量越高靠得就越近,也就更容易知道原子核结构)。穆斯保尔效应主要是能够非常方便地实现原子核的共振(原理是因为可以去除原子核的反冲),起作用是测量两体系间特别微小的能量差,具有精度高,成本低的优点,应用非常广泛。涉及物理学许多分支以及化学、生物学、地质学、冶金学、材料科学、环境科学和考古学等方面,形成了一门新的跨学科领域——穆斯保尔谱学。

**1962 年诺贝尔物理学奖**。开创了凝集态物质特别是液氦理论。获奖者:朗道(苏联)。评价:朗道是苏联历史上最著名的物理学家,其成就涉及物理学的各个方面。科学史上有所谓“朗道十戒”来概括朗道一生的成就。其实除此之外,朗道还有不少其他贡献,如恒星坍缩的质量,就是恒星到什么质量变成白矮星,什么质量变中子星,以及什么质量变黑洞。而液氦理论是十戒之一,当液氦降温降到一定的程度时密度会反常变小而不是变大(这很容易理解,一般物体都是热胀冷缩)。而朗道开创了凝集态物质理论对此的解释有突出贡献。朗道天赋过人,为人高傲,目空一切,所以遭到了不少人的嫉恨。在苏联大清洗中也难逃厄运,好在有贵人相助(莫斯科物理问题研究所所长卡皮查和他的导师玻尔)。由于他过于自负使得最早发现宇称不守恒的沙皮罗与诺贝尔奖失之交臂,不过当他得知此事后在杨李二位的基础上发展出弱相互作用下宇称电荷联合守恒(CP 守恒,十戒之一)。而且朗道极为关心教育,无论是从中学基础物理教育到最前沿的物理专业朗道都留下了有深刻影响的著作。他对学生要求极为严格,他的学生要做大量的习题,毕业之前还要通过朗道难度极大的考试。他和列夫谢兹编写的物理学教程深度和难度都很大,被奉为 20 世纪物理学的经典著作。朗道的学生在进行科研工作之前都要通读此书,学生戏称其为“朗道势垒”。不过就在朗道学术上最辉煌的时候,一场意外的车祸结束了他的学术生涯。而诺贝尔评审委员会为避免遗憾破例当年在莫斯科为朗道颁奖,可以说朗道是 20 世纪最有个性的科学家,是人类历史上最后一个全能物理学家。

**1963 年诺贝尔物理学奖**。发现基本粒子的对称性以及原子核中支配质子与中子相互作用的原理。获奖者:E. P. 威格纳(美国)。从事原子核壳层模型理论的研究。获奖者:M. G. 迈耶(美国)、J. H. D. 延森(德国)。评价:虽说杨振宁和李政道发现了宇称不守恒,但那只是在落弱相互作用下而言,在其他情况下宇称守恒还是物质普遍遵循的定律,而发现他的 E. P. 威格纳也就理所当然的获奖。而后二人在解释原子核模型上成就突出,阐明了当核子(质子和中子)数目为幻数(2、8、20、28、50、82 和 126)时为什么原子核结构比较稳定。

**1964 年诺贝尔物理学奖**。发明微波射器和激光器,并从事量子电子学方面的基础研究。获奖者:C. H. 汤斯(美国)、N. G. 巴索夫、A. M. 普罗霍罗夫(苏联)。评价:激光对人类的贡献想必不用多说,通信,医学,材料加工,军事都得靠它。而且如果没有激光很多光学实验还未必能方便地做出来。正因为它有广泛的用途从提出概念和原理后世界上许多实验室都被卷入了一场激烈的研制竞赛,看谁能成功制造并运转世界上第一台激光器。1960 年,美国物理学家西奥多·梅曼在佛罗里达州迈阿密的研究实验室里,勉强赢得了这场世界范围内的研制竞赛。1960 年 12 月,出生于伊朗的美国科学家贾万率人终于成功地制造并运转了全世界第一台气体激光器——氦氖激光器。1962 年,有三组科学家几乎同时发明了

半导体激光器。1966年,科学家们又研制成了波长可在一段范围内连续调节的有机染料激光器。此外,还有输出能量大、功率高,而且不依赖电网的化学激光器等纷纷问世。

**1965年诺贝尔物理学奖**。在量子电动力学方面进行对基本粒子物理学具有深刻影响的基础研究。获奖者:朝永振一郎(日本)、J. S. 施温格、R. P. 费曼(美国)。评价:科学家们根据量子力学理论推广到麦克斯韦方程中去发现会出现一些质量和电荷无穷大的点,显然这不并不是我们观察到的。而此三人的贡献就是想办法除去这些点并使之与实验结果符合。费曼被认为是美国出生的最伟大的理论物理学家,其成就对现在的理论物理影响深远。此人有着强烈的好奇心,为人风趣,别具一格被称为"即是滑稽演员又是科学家。"而且并对历史学也有突出贡献,玛雅象形文字的破译就有他的功劳。

**1966年诺贝尔物理学奖**。发现和发展研究原子中赫兹共振的光学方法。获奖者:卡斯特勒(法国)。评价:卡斯特勒的发现可以大大增加了探测磁共振信号的灵敏度,使人们多了一个研究原子能级结构的精密手段。不过具体解释起来非常麻烦,正如瑞典皇家科学院的一位发言人说:"要将此项工作解释给那些非科学家听几乎是不可能的。大概是通过施加一个作用来改变原子跃迁的能级分布,通过前后跃迁产生光的差别来研究原子的结构。"此外,还有以他名字命名的卡斯特勒奖主要是用来奖励固态物理学、原子与分子物理学研究中取得创造性贡献的人士。只有发展中国家的国民,并在本国生活与工作,年龄在40岁以下的学者(这条和数学诺贝尔菲尔兹奖差不多),才有申请这一奖励的资格。

**1967年诺贝尔物理学奖**。核反应理论的贡献,特别是发现恒星能源。获奖者:H. A. 贝蒂(美国)。评价:一直以来人们都对太阳的能源产生了极大的兴趣,因为它是如此持久,毕竟一个同等大小的煤球是只能支持几千年的。后来随着相对论的提出和聚变反应的发现,人们认为太阳的能源来源于聚变反应,但具体是如何产生的还不清楚。贝蒂就提出了氢—氢循环和氢—碳循环的说法解释了谜团(虽然后面还有其他形式但主要是这两种),而且根据此理论还可以解释我们这个世界上各种元素是如何起源的。

**1968年诺贝尔物理学奖**。通过发展液态氢气泡和数据分析技术,从而发现许多共振态。获奖者:L. W. 阿尔瓦雷斯(美国)。评价:高能(6.2GeV)质子同步稳相加速器首次运转时,识别到某些粒子。由于这些粒子的寿命远比理论模型所预言的产生时间长得多,因而人们将它们称为"奇异粒子"。但当时的仪器又无法观察它们。到了格拉泽发现气泡室后,事情有了转机。阿尔瓦雷斯对其发展了气泡室,并取得重要进展。

**1969年诺贝尔物理学奖**。对基本粒子的分类及其相互作用方面的卓越贡献。获奖者:M. 盖尔曼(美国)。评价:这次物理学奖依然给了在基本粒子上取得贡献的人,看来人类对微观世界的追求还真是执着。盖尔曼的主要贡献是发明了一种既能解释多重态中粒子的性质,又能预言新粒子存在的理论,这有点类似于门捷列夫在构造元素周期表的理论。而且在此基础上提出了夸克模型。这是现在人类解释原子核的模型。夸克是一种海鸟的叫声,表示他在提出这个观点时也觉得这个想法很荒谬。

**1970年诺贝尔物理学奖**。从事铁磁和反铁磁方面的研究。获奖者:L. 内尔(法国)从事磁流体力学方面的基础研究。获奖者:H. 阿尔文(瑞典)。评价:阿尔文关于磁流体动力学的研究对于受控热核反应的发展,超音速飞行,为外空推进器提供动力以及飞行器重新进入地球大气圈时的制动都起着非常重要的作用。而阿尔文提出的反铁磁性在电话学、磁带涂层、计算机的记忆磁芯和低损耗的高频率技术中都极其有用。

## 思考与练习

1. 相干必要条件中“相位差恒定”的物理意义是什么？
2. 普通热光源发光特点是什么？
3. 相干时间与相干长度的物理意义是什么？
4. 时间相干性与空间相干性没有联系吗？
5. 光程差变化时，多波长干涉与双波长干涉干涉花样的变化有何不同？
6. 什么是分波阵面干涉？什么是分振幅干涉？
7. 什么是等倾干涉？等倾干涉中的“等倾”的物理意义是什么？
8. 什么是等厚干涉？等厚干涉中的“等厚”的物理意义是什么？
9. 为什么多光束干涉才可以获得细锐度很好的干涉条纹？
10. 自由光谱程的物理意义是什么？

# 第9章

# 光的衍射与全息照相

在人们的日常生活中，偶尔能观察到光的衍射现象，例如，当你眯缝着眼睛看远处一盏发光的路灯，就会看到它向上下方发出长长的光芒，这就是光在视网膜上的衍射图像。所谓光的衍射，就是光波遇到小障碍物或小孔时，绕过障碍物进入几何阴影区继续传播，并在障碍物后的观察屏上呈现光强的不均匀分布的现象。但是由于光的波长很短，并且普通的光源又是相干性很差的面光源(关于相干性的详细讨论，请参阅第8章)，因此，在日常生活中光的衍射现象很少被人观察到，相反，光的直线传播却给人们留下了深刻印象。

实际上，光作为一种高频电磁波，在传播过程中无论是在自由空间，还是在有界空间，均有衍射现象伴随。从严格意义上讲，衍射是波动在传播过程中其波面受到限制的必然结果，而不单纯是一种边缘效应。衍射理论是现代变换光学的理论基础。任何一个光学系统，都是一个有界的光波场传输系统，因而都存在衍射现象。

在光学的发展历史上，光的衍射现象是光具有波动特性的有力证据；在光学的应用方面，光的衍射是限制几何光学仪器分辨本领的根本因素。激光诞生以后，利用光的衍射现象，使全息照相得以实现。本章从光的衍射的角度，对光的衍射现象的物理本质、实验方法、数学描述方法以及应用等方面做比较系统的讨论。

## 9.1 从惠更斯原理到惠更斯-菲涅耳原理

纵观物理学的发展，围绕着物质本性的探究，波动性和粒子性作为其两个侧面一直是物理学研究的主旋律。作为波动说的基石，惠更斯-菲涅耳原理自然显得特别重要。透彻地理解它无论对物质的本质理解还是对一种思维方法的掌握都有很大的帮助。

### 9.1.1 惠更斯原理的提出

1679年，惠更斯向法国科学院提交的报告中提出了他著名的光波动理论：“光波发射时，传播光的每一物质粒子不只把运动传给前面的邻近粒子(与原始粒子和光源位于同一直线上)，而且还应该传给周围所有其他和自己接触并阻碍自己运动的粒子，因此，在每一粒子周围就产生以此粒子为中心的波，即子波。”引入子波(次波)概念后，则通过几何方法寻找包络面，就能确定各种情况下波的传播方向。惠更斯为了说明波在空间各点的逐步传播的机制，曾经提出这样的设想，现今称之为惠更斯原理。

惠更斯原理可以表述如下：

媒质中的波传到的各点，都可看作开始发射子波(次级波)的子波源(是点波源)，在以后的任一时刻，这些子波面的包络面就是实际的波在该时刻的波前。

如图9-1-1所示，图中媒质均匀各向同性，各子波都是以波速 $u$ 向外扩展的球面波。若已知 $t$ 时刻的波面就可以知道 $t+\Delta t$ 时刻的波面，从而确定波的传播方向。

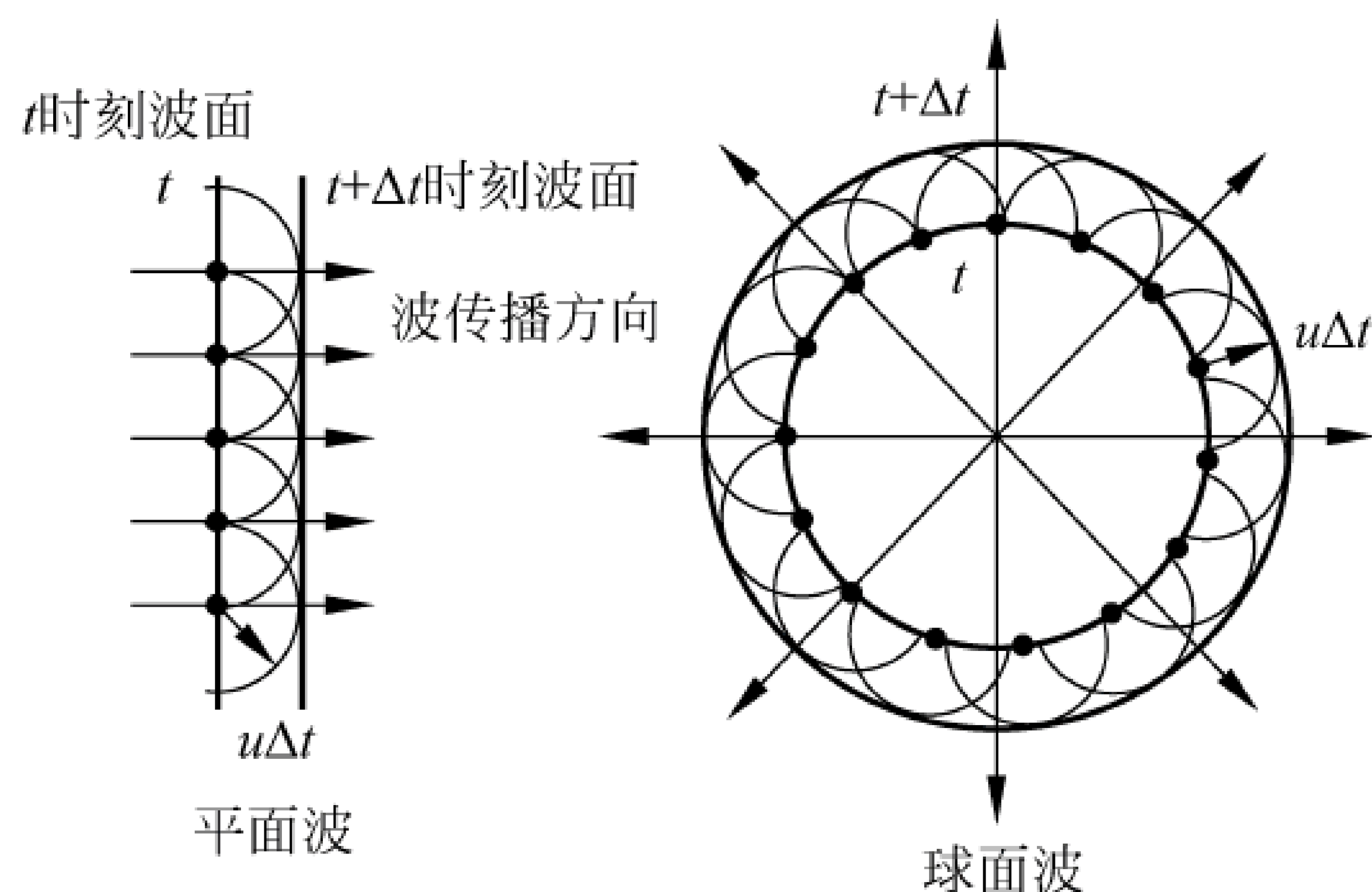



图 9-1-1　惠更斯原理

## 9.1.2 惠更斯原理的应用

根据惠更斯原理，可以解释光的直线传播、反射、折射和双折射等现象，还可以预示光的衍射现象的存在。

### 1. 光的折射

如图 9-1-2 所示，一束平面光波倾斜入射到两种介质的分界面上，并在该界面上发生折射现象，取 $i_1$ 为入射角，$i_2$ 为折射角，两种介质的折射率分别为 $n_1$ 和 $n_2$，光波在相应介质中的速度分别为 $u_1$ 和 $u_2$。

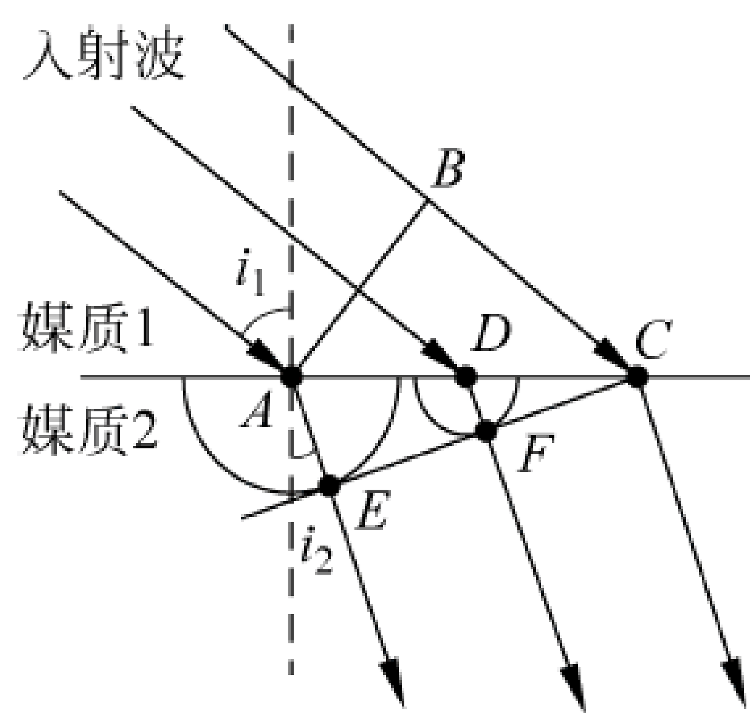



图 9-1-2　折射定律

设 $t_1$ 时刻入射波到 $A$ 点，波前为 $AB$；$t_2$时刻入射波到 $C$ 点，则

$$\overline{BC}=u_1(t_2-t_1) \tag{9-1-1}$$

$A$、$D$、$C$ 各点向媒质 2 所发子波在 $t_2$时刻（$C$ 到界面的时刻）的子波面如图 9-1-2 所示，于是有

$$\overline{AE}=u_2(t_2-t_1) \tag{9-1-2}$$

子波波面的包络面（图中 $EFC$）就是媒质 2 中的波前。通过入射点和切点（包络面与子波面的切点）的直线就是折射波的传播方向。由图有，光的折射定律

$$\frac{\sin i_1}{\sin i_2}=\frac{u_1}{u_2}=\frac{n_1}{n_2} \tag{9-1-3}$$

### 2. 光的衍射

同样，也可以由惠更斯原理对光的衍射现象做出定性解释。

如图 9-1-3(a)所示，考察一平面波 $WW'$通过宽度为 $a$ 的开孔 $AA'$的情况。开孔 $AA'$限制平面波 $WW'$只允许宽度为 $a$ 的一段波阵面通过，开孔平面上的每一点都可视为新的光源，传出次级波；这些次级波的包络面在中间部分是平面，在边缘处是弯曲的。即在开孔的边缘处不沿原光波方向进行，因而可以预料在几何影内的光强度不为零，它表明有衍射现象，是光的波动性的表现。在几何观点上，自点（或线）光源发出的光波，当其通过圆孔、狭缝、直边或其他任意形状的孔或障碍物面到达幕上时，在幕上应该呈现明晰的几何影，影内完全没有光，影外有均匀的光强度分布；但在实际情况，在圆孔、狭缝、直边或其他任意形状的孔或障碍物都很小的情况下，由于它们限制光波的波阵面，结果有光进入影内并且在影外的光强分布也不均匀，这是光的直线传播定律所不能解释的，这种现象称为光的衍射。与上面所论述的相比较，可以得知惠更斯原理可以预示光的衍射现象的存在。而且，孔或障碍物

的尺寸越小,阴影区的光强分布展开得越宽,衍射现象越明显(见图 9-1-3(b))。

惠更斯原理的应用范围很广,机械波和电磁波都可以应用。惠更斯就是在研究光波的传播时提出此原理的。当时惠更斯认为光是波动的,而牛顿则认为光是微粒流,由于牛顿当时在科学界的影响力很大,所以,惠更斯原理没有得到公认。直到 100 年之后,托马斯·杨于 1801 年成功研究了光波双缝干涉实验,惠更斯原理才被公认。

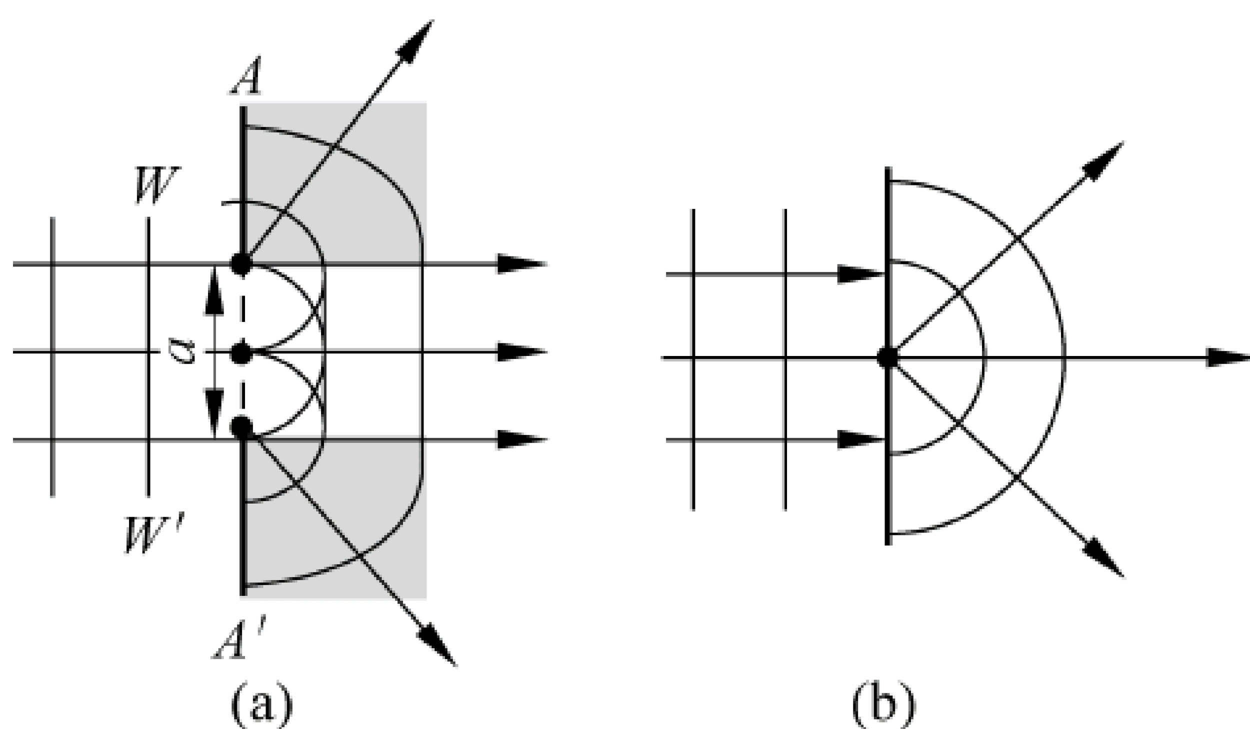



图 9-1-3 光的衍射

### 9.1.3 惠更斯原理的局限性

惠更斯原理不能定量地研究波,也不能解释子波为何只向前发射,而不向后发射,特别是不能说明波的衍射现象中能量是怎么分布的,不能讨论波的强度问题。也即是惠更斯原理有助于确定光波的传播方向,而不能确定沿不同方向传播的振动的振幅。这些问题在 1818 年菲涅耳进一步提出惠更斯-菲涅耳原理后才得到解决。

### 9.1.4 惠更斯-菲涅耳原理的提出

由于惠更斯原理的次波假设不涉及波的时空周期特性——周期、波长、振幅和相位,因而不能说明在障碍物边缘波的传播方面偏离直线的现象。事实上,光的衍射现象要细微得多,还有明暗相间的条纹出现,表明各点的振幅大小不等。因此,必须定量计算光所到达的空间范围内任何一点的振幅,才能准确地解释衍射现象。菲涅耳(Fresnel Augustin Jean,1788—1827)根据惠更斯的"次波"假设,仔细分析了次波性质,发现次波满足相干的必要条件,所以,他补充了描述次波的基本特征量——相位和振幅的定量表示式,并增加了"次波相干叠加"的原理,使之发展成为惠更斯-菲涅耳原理。惠更斯-菲涅耳原理成为定量研究波的相干叠加的理论基础。

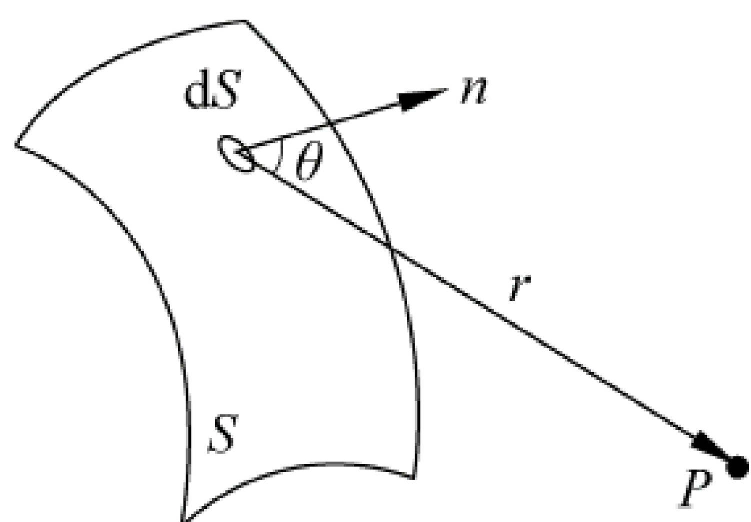



图 9-1-4 惠更斯-菲涅耳原理图

惠更斯-菲涅耳原理(Huygens-Fresnel principle)内容的表述如下(见图 9-1-4):

面积元 $dS$ 所发出的各次波的振幅和位相满足下面四个假设:

(1) 在波动理论中,波面是一个等位相面。因而可以认为 $dS$ 面上各点所发出的所有次波都有相同的初位相(可令 $\varphi=0$)。

(2) 次波在 $P$ 点处所引起的振动的振幅与 $r$ 成反比。这相当于表明次波是球面波。

(3) 从面元 $dS$ 所发次波在 $P$ 处的振幅正比于 $dS$ 的面积,且与倾角 $\theta$ 有关,其中 $\theta$ 为 $dS$ 的法线 $n$ 与 $dS$ 到 $P$ 点的连线 $r$ 之间的夹角,即从 $dS$ 发出的次波到达 $P$ 点时的振幅随 $\theta$ 的增大而减小,用倾斜因数 $K(\theta)$ 来描述。

(4) 次波在 $P$ 点处的位相,由光程 $\Delta=nr\left(\text{对应的相位差为 } \varphi=\frac{2\pi}{\lambda}\Delta\right)$决定。

根据以上的假设,可知面积元 $dS$ 发出的次波在 $P$ 点的合振动可表示为

$$dE \propto \frac{K(\theta)}{r}\cos(kr-\omega t)dS \tag{9-1-4}$$

或

$$dE = C\frac{K(\theta)}{r}\cos(kr - \omega t)dS \tag{9-1-5}$$

如果波面上各点的振幅有一定的分布，则面元 $dS$ 发出次波到达 $P$ 点的振幅与该面元上的振幅成正比，若分布函数为 $A(Q)$，则波面在 $P$ 点所产生的振动为

$$dE = C\frac{K(\theta)A(Q)}{r}\cos(kr - \omega t)dS \tag{9-1-6}$$

如果将波面上所有面积元在 $P$ 点的作用加起来即可求得波面 $S$ 在 $P$ 点所产生的合振动：

$$E = \int_S dE = C\int\frac{K(\theta)A(Q)}{r}\cos(kr - \omega t)dS \tag{9-1-7}$$

也可以写成复数形式：

$$E = C\int\frac{K(\theta)A(Q)}{r}e^{i(kr-\omega t)}dS \tag{9-1-8}$$

式(9-1-8)称为菲涅耳衍射积分，有的文献称之为基尔霍夫公式，一般来说计算此积分式是相当复杂的，但在波面对于通过 $P$ 点的波面法线具有旋转对称性的情况下，积分就比较简单，可用代数加法或矢量加法来代替积分。

惠更斯-菲涅耳原理抓住了波动的两个基本性质，为光场提供了一幅比较简明而又生动形象的物体图像。我们知道，凡是波动都具有两个基本属性，一方面它是扰动的传播，一点的扰动能够引起其他点的扰动，各点之间的扰动是有联系的；另一方面，它具有时空周期性，能够相干叠加。光场也是一样，由于光场中任何一点具有变化的光矢量，在它周围就会产生变化的磁场，变化磁场又在它的周围产生变化的电场。在这种意义上，我们可以认为光场中的每一点都可以看作一个“次光源”。而且光场中各点的光振动又不是孤立的，它们通过“次波”的相干叠加，互相影响，形成整体。

### 9.1.5　惠更斯-菲涅耳原理的应用

借助于惠更斯-菲涅耳原理可以解释和描述光束通过各种形状的障碍物所产生的衍射现象。我们知道两束相干光互相重叠时会发生干涉，产生干涉条纹。但是在实验室中用光源照射一个狭缝、一根金属丝、一个小圆孔或一个小圆屏，也会看到明暗条纹，这是光的衍射现象。下面用惠更斯-菲涅耳原理来解释这一类现象。

一束光的波阵面到达狭缝的时候，按照惠更斯原理，此狭缝的波阵面上各点都可以看成是发射子波的波源，这些子波可以绕射到狭缝的前方和两侧。但是，这些子波为什么会形成明暗相间的条纹呢？

为了说明上述现象，菲涅耳用干涉原理给惠更斯原理作了补充，他认为惠更斯原理中的子波是从同一波阵面发出的，因而在重叠区域满足频率相同和相位差恒定的条件，又因为子波的传输范围不太大，因而振动分量相互平行自然会满足，重叠在一起的时候就会发生干涉。因此，在有些位置上，子波因为干涉加强而形成亮纹；有些位置，子波因为干涉相消而形成暗纹。此现象称为光的衍射现象或者绕射现象，它是波动现象的重要特征之一。其他电磁波和机械波也会发生衍射现象，也都可以应用上述的惠更斯-菲涅耳原理去分析计算。

按照菲涅耳的观点，衍射现象是无数子波的干涉现象。因此衍射与干涉之间不存在实质性的物理区别。但是在分析、计算衍射现象时，要注意它是无数子波的叠加。

此外，惠更斯-菲涅耳原理还可以讨论不同形状的孔和障碍物产生的衍射图样的光强

分布。

### 9.1.6 衍射现象的分类

衍射系统由光源、衍射屏和接收屏组成。通常按它们相互间的距离大小,将衍射分为两类:一是菲涅耳衍射(Fresnel diffraction)或称近场衍射,即光源和观察屏(或两者之一)离衍射屏距离有限远(图 9-1-5(a))。这时入射光和衍射光(或两者之一)不是平行光,波面的曲率不可忽略。二是夫琅禾费衍射(Fraunhofer diffraction)或称远场衍射,即光源和观察屏距衍射屏均为无限远(图 9-1-5(b)),因此入射光和衍射光都是平行光,波面的曲率可以忽略。实际的装置必须利用两个会聚透镜,让点光源和接收屏分别位于两个透镜的前、后焦距位置,以保证入射光和衍射光都是平行光,故夫琅禾费衍射又称为平行光衍射。

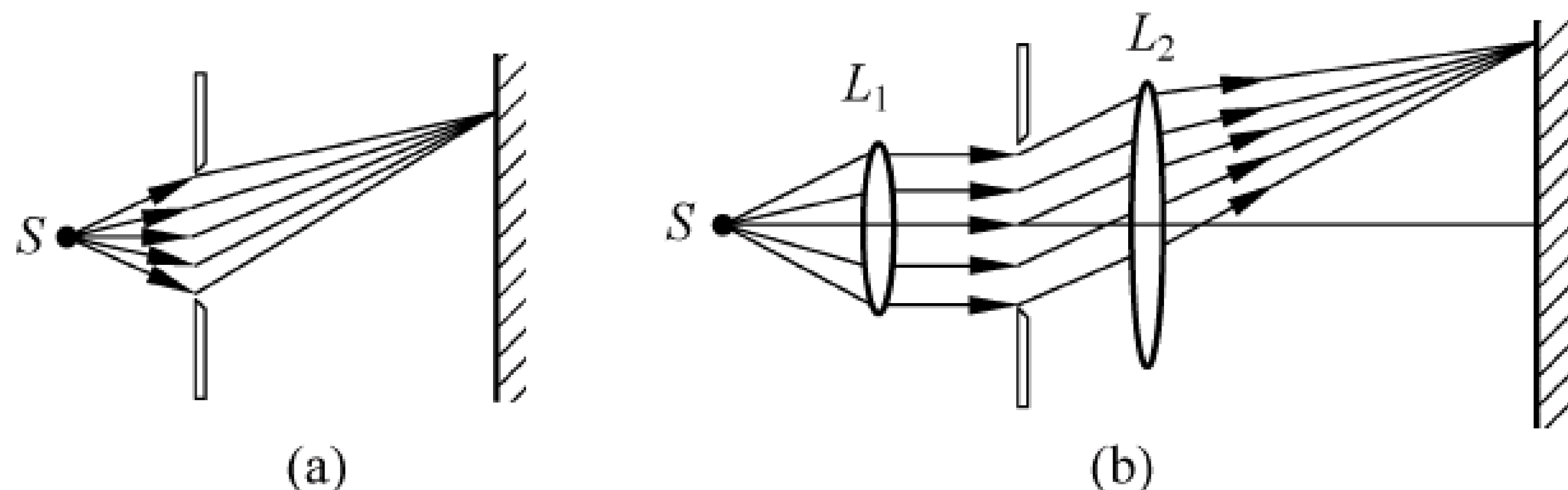



图 9-1-5 衍射的分类

最初这种分类原则是从理论计算上考虑的。夫琅禾费衍射场的计算比较容易,可以用衍射积分公式进行比较严格的计算;而菲涅耳衍射场的计算较为复杂,不易用积分公式直接计算,通常用菲涅耳“半波带法”和“矢量叠加法”作定性的或半定量的近似计算,以便发现衍射场的基本特征。但近来发展起来的傅里叶变换光学,赋予了夫琅禾费衍射以新的重要意义,给了它新的定义。这也使得这样的分类原则有了新的含义。

## 9.2 费马原理与几何光学的等光程性

我们知道,光在不同介质中传播时,其频率是不变的,但是其速度不同,这样,在相同的时间内,光在不同介质中传播的几何路程就不相同。为了准确描述光在不同介质中传播时相位的变化,必须对光在不同介质中传播的几何路程进行比较,由此引入光程的概念。

### 9.2.1 光程

在均匀介质中,光程 $L$ 为光在介质中通过的几何路径 $S$ 与该介质折射率 $n$ 的乘积,即

$$L = nl \tag{9-2-1}$$

如图 9-2-1 所示,如果光线从 $A$ 点出发,中间经过 $N$ 种均匀介质而到达 $B$ 点,各种介质的折射率分别为 $n_1, n_2, \cdots, n_i, \cdots, n_k$,光线在其间传播的几何路径的长度分别为 $\Delta l_1, \Delta l_2, \cdots, \Delta l_i, \cdots, \Delta l_k$,传播速度分别为 $v_1, v_2, \cdots, v_i, \cdots, v_k$,则总光程 $L$ 为

$$L = n_1\Delta l_1 + n_2\Delta l_2 + \cdots + n_i\Delta l_i + \cdots + n_k\Delta l_k = \sum_{i=1}^{k} n_i\Delta l_i \tag{9-2-2}$$

若光线在每个介质区域的几何路径长度很小,或假设介质折射率连续变化(图 9-2-2),则式(9-2-2)的求和可由下面的积分代替:

$$L=\int_A^B n\,\mathrm{d}l \tag{9-2-3}$$

沿光线方向的元路程 $\mathrm{d}l>0$，逆光线方向的元路程 $\mathrm{d}l<0$。

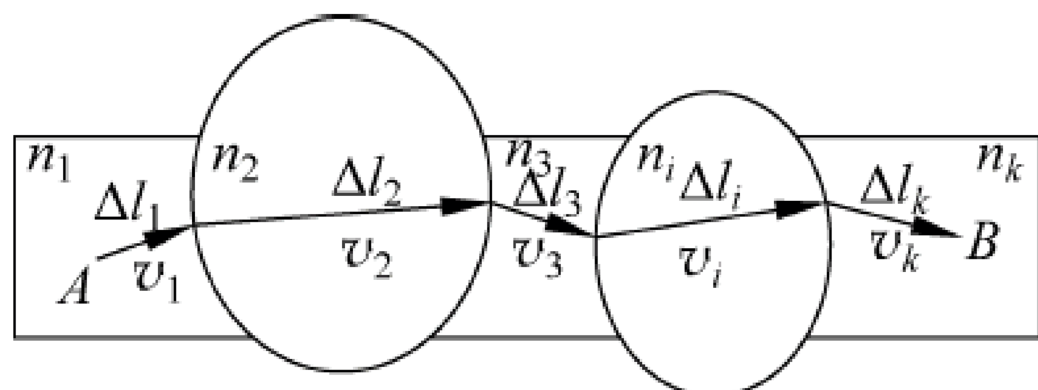



图 9-2-1　光程(一)

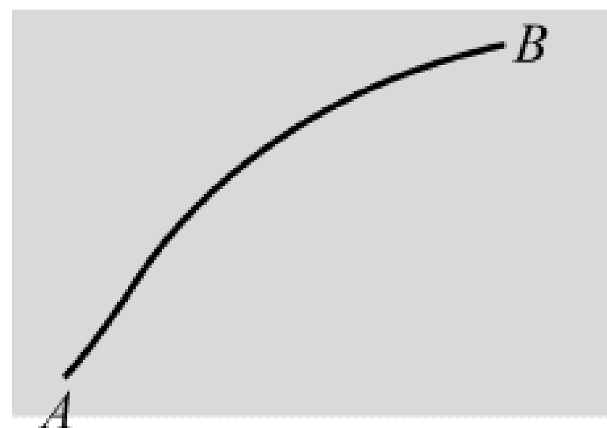



图 9-2-2　光程(二)

光从 $A$ 点到 $B$ 点的光程 $L$ 和所需时间 $t$ 之间有正比关系。为证明这一点，令 $v$ 表示光经过元路程 $\mathrm{d}l$ 时的速率，则

$$L=\int_A^B n\,\mathrm{d}l=c\int_A^B \frac{\mathrm{d}l}{v}=ct \tag{9-2-4}$$

式(9-2-4)也可以理解为：光程 $L$ 的数值等于在 $t$ 时间内，光在真空中能传播的几何路程。借助光程概念可将各种介质中走过的路程折算为真空中的路程，便于比较光在不同介质中用时的长短。

## 9.2.2　费马原理

自然界一切物质都在运动和变化之中，其变化和运动之趋势，一般是达到一种平衡状态，即极值状态。作为自然客体的一种，光的传播也必须服从一种极值规律，即费马原理(Fermat principle)。

**1. 费马原理的内容**

1657 年费马概括了光线传播的实验定律，提出了一条统一的原理：光从空间的一点 $A$ 到另一点 $B$ 的实际路径，与任何其他可能的邻近路径相比，其光程(或费时)为极值。简言之，光沿着光程(或费时)为极大、极小或者常数的路径传播，即在光线的实际路径上，光程(或费时)的变分为 0，有

$$\delta L=\delta\int_A^B n\,\mathrm{d}l=0 \tag{9-2-5}$$

或

$$\delta t=\frac{1}{c}\delta\int_A^B n\,\mathrm{d}l=0 \tag{9-2-5$'$}$$

如图 9-2-3 所示，如果 $ACB$ 代表光线的实际路径，光线 $ACB$ 的光程(或者说所需的时间)与邻近的任何可能路径 $AC'B$ 相比为极值(极大值、极小值或常数)。

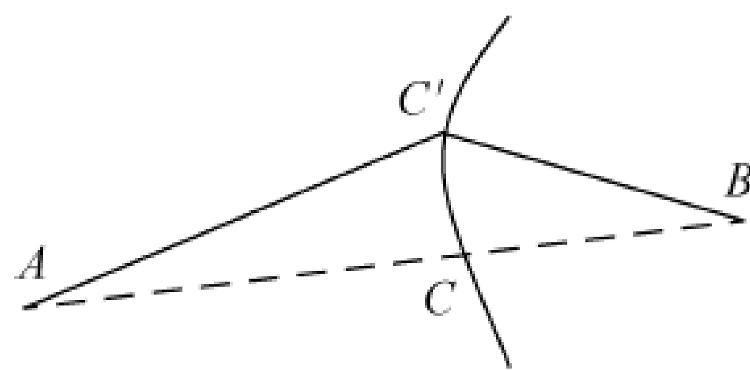



图 9-2-3　费马原理

几何光学的三个基本定律都可以用费马原理概括起来。因为费马原理不但可以描述光在均匀介质中的传播情况，还可以描述光在不均匀介质中的传播情况。

**2. 费马原理的验证**

如图 9-2-4 所示，可以用椭圆球面的内反射来验证费马原理。根据椭圆的几何性质，椭圆上任何一点到两个焦点的连线的夹角平分线就是过该点的法线，且两线段长度之和为常数。这说明对于内表面为镜面的椭球，从其焦点 $F_1$ 发出的光线无论经椭球内表面

的哪一点反射，反射光线都会以相等的光程到达椭球的另一个焦点 $F_2$，即光程取恒定值(常数)(图 9-2-4(a))。其实这也是极值，不妨设想有一个凹面反射镜与椭球面内切(图 9-2-4(b))，则自椭球面焦点 $F_1$ 发出的所有到达该凹面镜上的光线中，只有过切点 $M$ 的光线可以经该凹面镜反射到达椭球面的另一个焦点 $F_2$，显然该光线的光程为极大值。与此相反，若将一个凸面镜与椭球面外切(图 9-2-4(c))，则从椭球面焦点 $F_1$ 发出的所有到达该凸面镜上的光线中，只有过切点 $M$ 的光线才可以经该凸面镜反射到达椭球面的另一个焦点 $F_2$，此时光程为极小值。

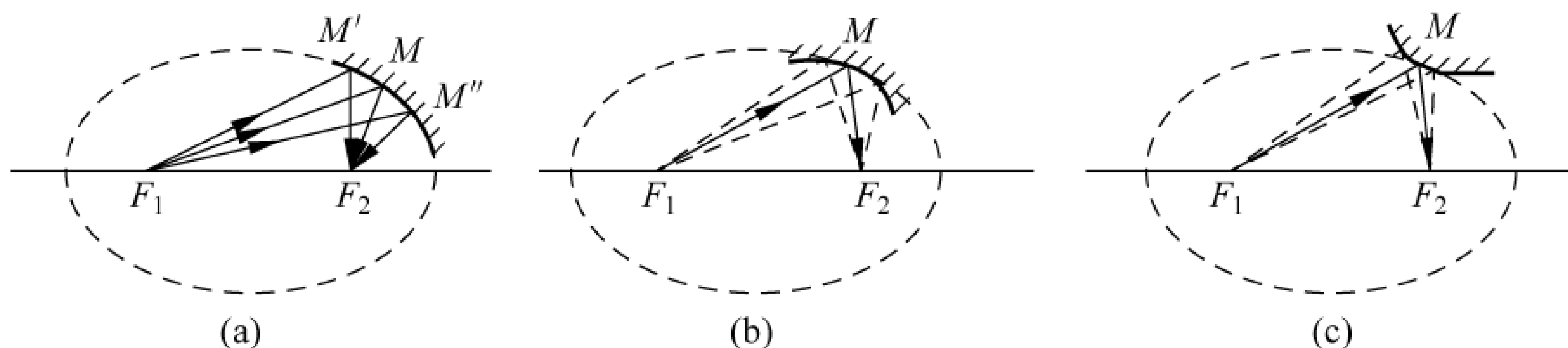



图 9-2-4　费马原理的验证

## 9.2.3　用费马原理推导几何光学规律

### 1. 光的直线传播和光路可逆

几何光学主要研究的是光的传播。而光的传播规律其中一条就是光在同种均匀介质中沿直线传播，可以由费马原理直接得出。两点之间的所有可能连线中，直线段最短，符合费马原理光程取极小值的要求，光的直线传播直接得证。

费马原理只涉及光线传播的路径，并未涉及光线的传播方向。若路径 $PQ$ 的光程为极值，则其逆路径 $PQ$ 的光程也为极值，由此便导出了光的可逆性原理。

### 2. 光的反射

如图 9-2-5 所示，考虑由 $Q$ 发出，经反射面 $\Sigma$ 上某点 $M$ 到达 $P$ 点的光线。$M$ 点位于何处，光程 $n(QM+MP)$ 才有极值呢？作点 $P$ 相对于反射面 $\Sigma$ 的对称点 $P'$，从 $Q$ 到 $P$ 任意可能路径 $QM'P$ 的长度与 $QM'P'$ 的长度相等。根据两点之间直线最短的公理，显然，直线段 $QMP'$ 是其中最短的一根，从而路径 $QMP$ 的长度最短。根据费马原理，$QMP$ 是光线的实际路径。由对称性分析不难看出，$\angle i'=\angle i$。

### 3. 光的折射

图 9-2-6 是用费马原理导出折射定律的示意图。平面 $\Sigma$ 上方的折射率为 $n_1$，下方的折射率为 $n_2$。$Q$、$P$ 两点所在的平面为 $\Pi$ 面。

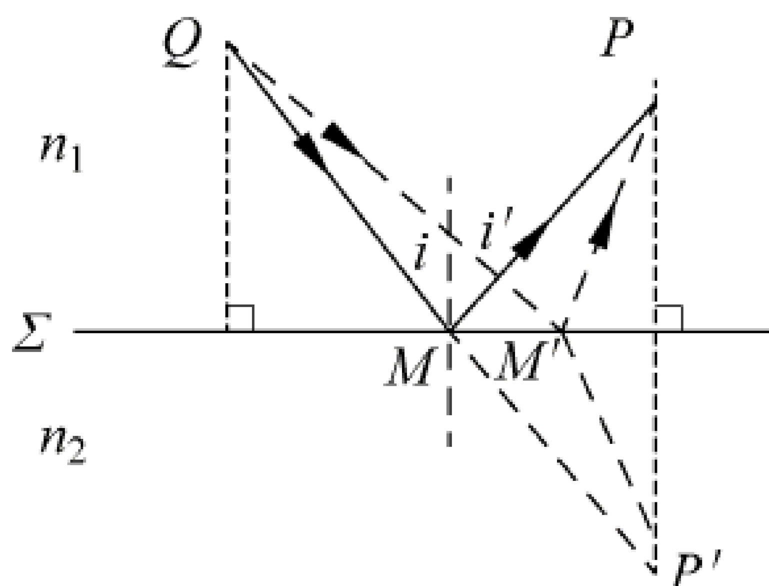



图 9-2-5　用费马原理推导反射定律

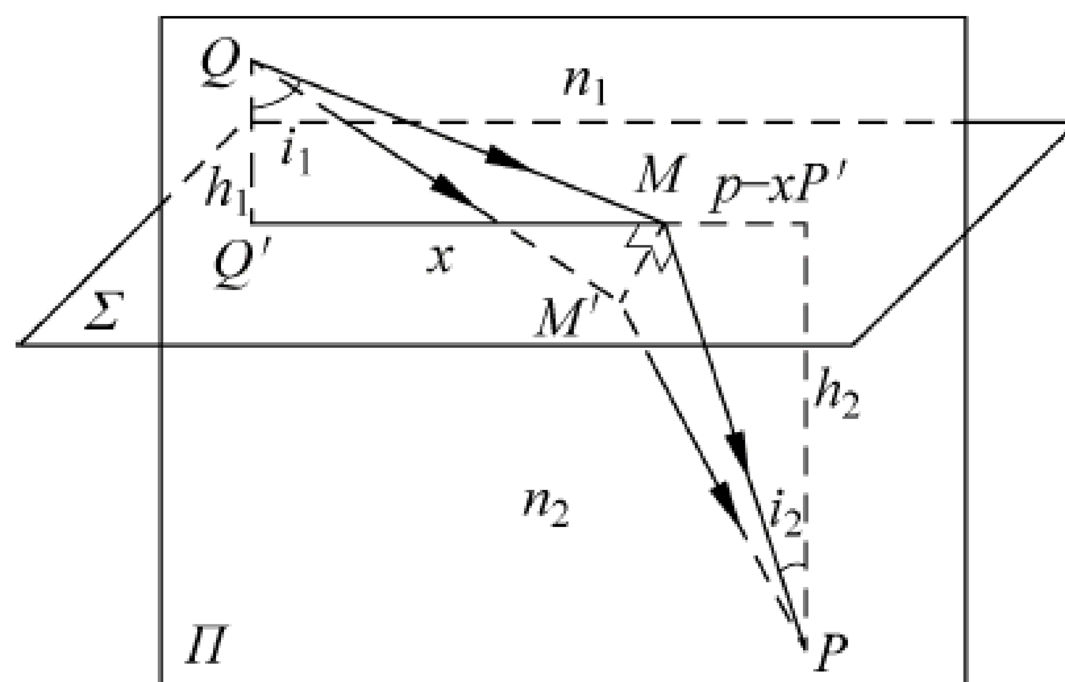



图 9-2-6　用费马原理推导折射定律

设 $Q$、$P$ 两点间光线路径 $QM'P$ 不在入射面 $\Pi$ 内，总可以从 $M'$ 点作平面 $\Sigma$ 垂直于入射面 $\Pi$，路径 $QMP$ 位于入射面 $\Pi$ 内，由于 $QM'>QM$，$M'P>MP$，故光程极小路径必须在入射面 $\Pi$ 内，其光程为

$$(QMP)=n_1QM+n_2MP=n_1\sqrt{h_1^2+x^2}+n_2\sqrt{h_2^2+(p-x)^2} \tag{9-2-6}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(QMP)=\frac{n_1x}{\sqrt{h_1^2+x^2}}-\frac{n_2(p-x)}{\sqrt{h_2^2+(p-x)^2}}=n_1\sin i_1-n_2\sin i_2 \tag{9-2-7}$$

由光程取极小值条件

$$\frac{\mathrm{d}}{\mathrm{d}x}(QMP)=0 \tag{9-2-8}$$

即得

$$n_1\sin i_1=n_2\sin i_2 \tag{9-2-9}$$

### 9.2.4　费马原理与几何光学的等光程性

图 9-2-7 中 $OO'$ 是平面镜，$S$ 和 $S'$ 是一对物像共轭点，可以证明：理想光学系统物像共轭点 $S$、$S'$ 间是等光程的。设物点 $S$ 所在一侧介质折射率为 $n$，$B$ 为镜面上任一点，则 $(SBS')=n\int_S^B \mathrm{d}l+n\int_B^{S'}\mathrm{d}l$。$BS'$ 段光程是反射线 $BC$ 延长线的光程，其折射率按反射线 $BC$ 所在介质的折射率 $n$ 计算，$\mathrm{d}l$ 沿 $BS'$ 走向与反射光 $BC$ 方向相反，故应取负值，即 $(SBS')=nSB+n(-BS')=0$。图 9-2-7 所示的平面镜成像只是理想光学成像系统等光程性的一个例子。下面我们再看另外两例。

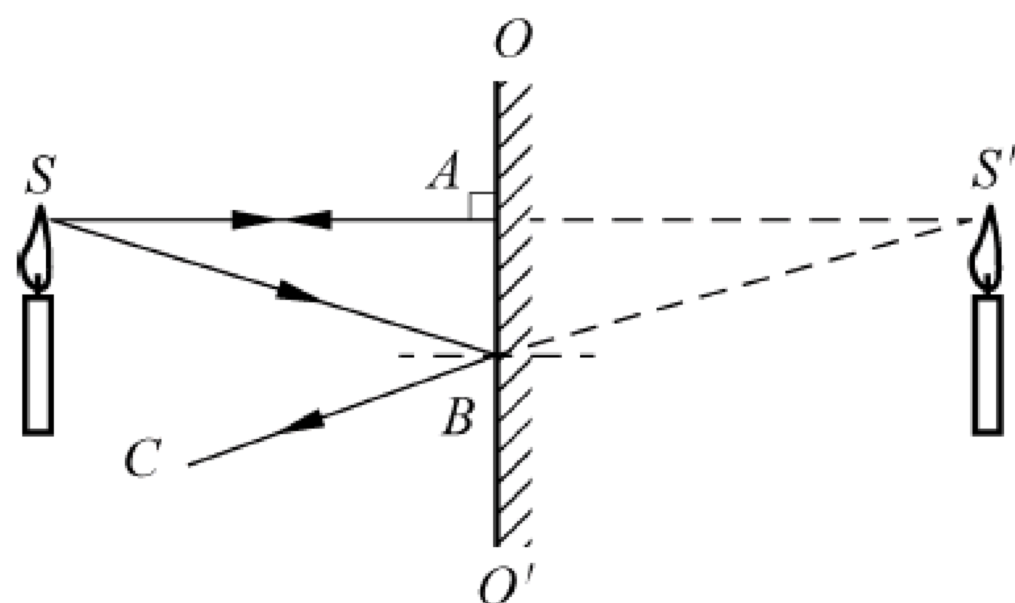


图 9-2-7　平面镜成像共轭点之间光线的等光程性

**例 1**　如图 9-2-8 所示，点物 $S$ 通过凸透镜成点像于 $S'$，由 $S$ 点发出的许多光线经透镜会聚于 $S'$，按费马原理，这些光线的光程必然彼此相等，不可能一条光线的光程比另一条的大。这就是物像之间的等光程性(或称等光程原理)。这一点很重要，在讨论衍射问题时常用到这一事实。

**例 2**　如图 9-2-9 所示，平行光垂直入射到凸透镜上，会聚于焦点 $F$。波面 $AB$ 垂直于入射平行光。试证明：图中的光线 $AF$、$EF$、$BF$ 的光程都是相等的。(提示：在图 9-2-8 中，点光源 $S$ 远离透镜时，入射光即近似为平行光)

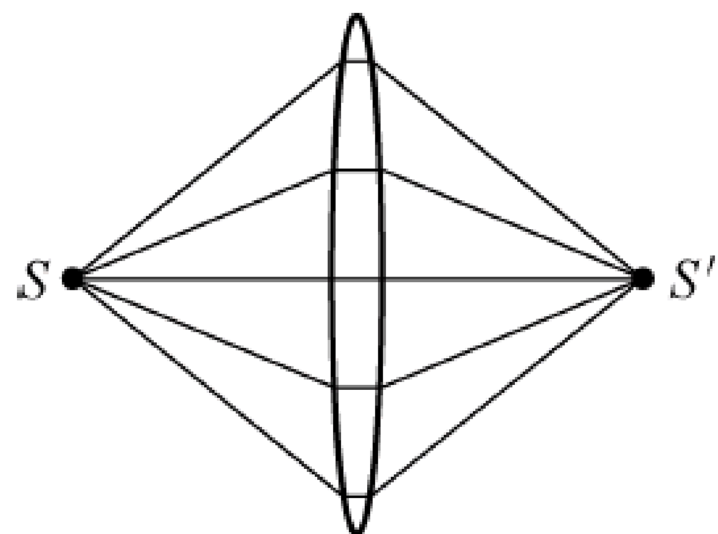


图 9-2-8　物像之间的等光程性

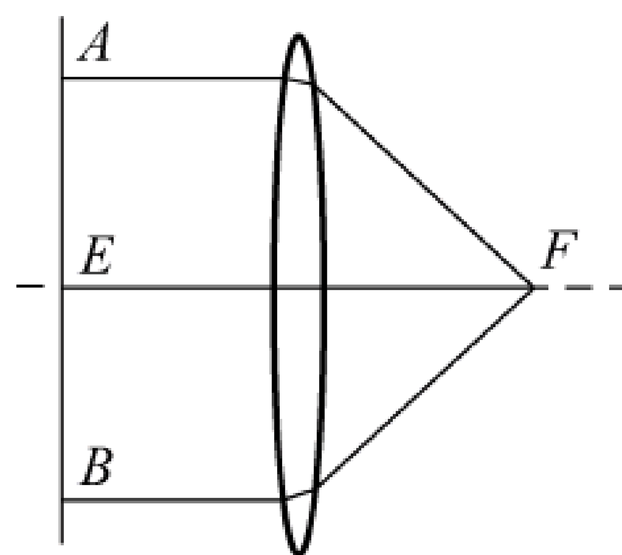


图 9-2-9　$AF$、$EF$、$BF$ 的等光程性

从上面的例子可以看出，所有能将点物成点像的理想光学成像系统，物像共轭点之间各光线都有等光程的特点。

### 9.2.5 费马原理的应用

从费马原理可以导出几何光学的全部定律,利用费马原理提供的寻找光束路径的基本方法,还可以为解决很多实际问题另辟蹊径,例如流传千年的“胡不归”问题。

古老的“胡不归”传说说的是:从前有一个身在 $A$ 地当学徒的小伙子,当他得悉在家乡 $B$ 地的年老父亲病危的消息后,便立即向掌柜告了假借了些钱启程赶路,由于思念心切,他挑选了全是沙砾地带的直线路径 $AB$,如图 9-2-10 所示,他认为走近路必定最省时,因此,他放弃了沿驿道 $AC$ 先走一程的想法。当他气喘吁吁地来到父亲跟前时,老人刚刚咽了气,小伙子不觉失声痛哭。邻舍闻声前来劝慰,有人告诉小伙子,老人在弥留之际,还不断喃喃地叨念“胡不归?胡不归?……”。并且深为怜惜地问道:“你为什么不向掌柜借用一下马车,沿驿道先走一程呢?”

由上述古老的传说,引起人们的思索,小伙子是否有可能提前抵达家门呢?倘若有可能,应该选择一条什么样的路线呢?这就是曾经风靡千年的“胡不归”问题。

在利用费马原理解决“胡不归”问题时,把小伙子看作光粒子,光粒子就是其物理模型,然后,根据光的折射定律建立其数学模型,可以巧妙地解决“胡不归”问题。

设想从 $A$ 点发出的一束光(代替小伙子)先与两媒质(驿道跟沙砾地带)界面 $AC$(即驿道)成一很小的角度入射到 $D$ 点,此时光速为 $v_1$,然后折射入第二种媒质(沙砾地带)到达 $B$ 点,此过程中光速为 $v_2$,假定此光束沿 $ADB$ 路线传播是符合折射定律的路径,过入射点 $D$ 作界面 $AC$ 的法线,得入射角 $\alpha\approx 90°$,折射角为 $\gamma$,如图 9-2-11(a)所示,根据折射定律有

$$n_{12}=\frac{\sin\alpha}{\sin\gamma}=\frac{v_1}{v_2} \tag{9-2-10}$$

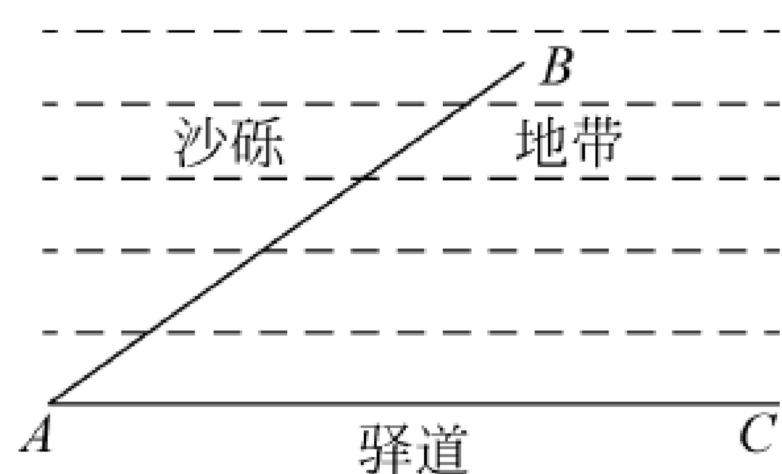


图 9-2-10 “胡不归”问题

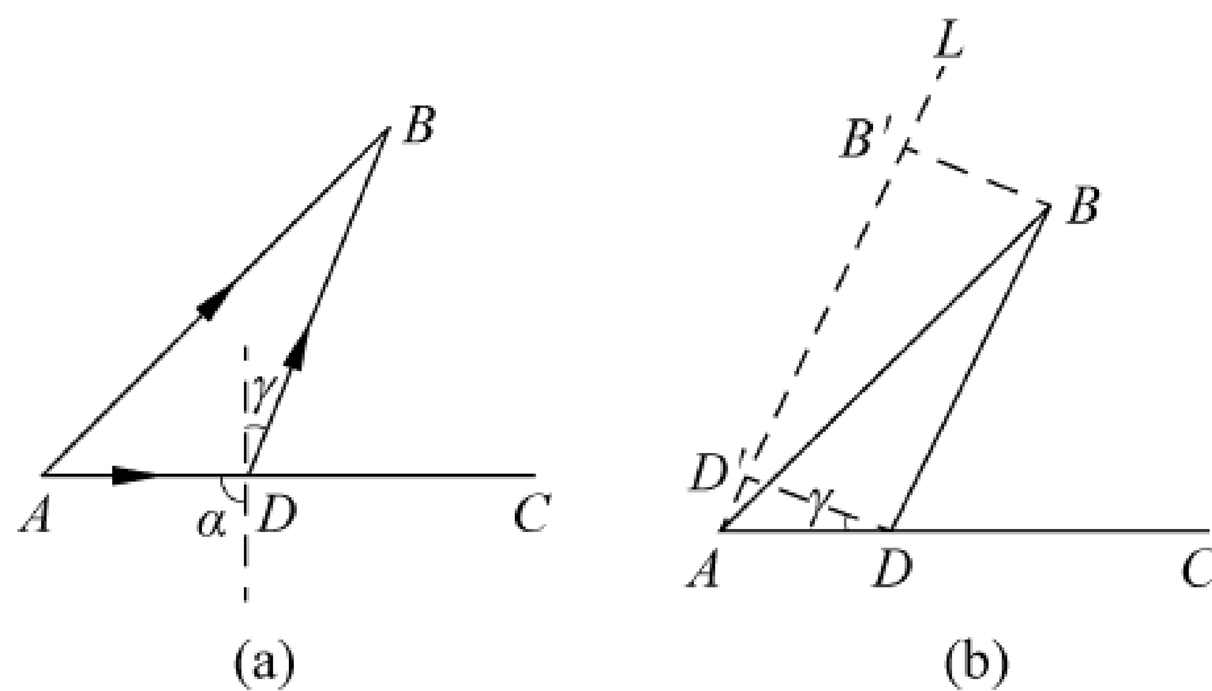


图 9-2-11 “胡不归”问题的模型

光束由 $A\to D$,$D\to B$,总共所需时间为

$$t_1=\frac{AD}{v_1}+\frac{DB}{v_2} \tag{9-2-11}$$

又若此束光从 $A$ 点发出,沿 $AB$ 直线,以速率 $v_2$ 传播(即小伙子所走的实际路线),其所需时间为

$$t_2=\frac{AB}{v_2} \tag{9-2-12}$$

为了比较 $t_1$ 与 $t_2$,先过 $A$ 点作 $DB$ 的平行线 $AL$,再过 $D$、$B$ 两点作 $AL$ 的垂线,分别交于 $D'$点和 $B'$点,如图 9-2-11(b)所示。显然在三角形$\triangle ADD'$中,$\angle ADD'=\gamma$,$\angle DD'A=\alpha$,由正弦定律可得

$$\frac{AD}{\sin\alpha}=\frac{AD'}{\sin\gamma} \tag{9-2-13}$$

由式(9-2-10)可得

$$\frac{AD}{AD'}=\frac{v_1}{v_2} \tag{9-2-14}$$

再将式(9-2-14)代入式(9-2-11)消去 $v_1$,并注意到 $DB=D'B'$,则

$$t_1=\frac{AD}{v_1}+\frac{DB}{v_2}=\frac{AD'}{v_2}+\frac{D'B'}{v_2}=\frac{AB'}{v_2} \tag{9-2-15}$$

比较式(9-2-12)和式(9-2-15),因在直角△$ABB'$中,$AB>AB'$,最后可得:$t_1<t_2$。上述最后结果表明:小伙子本来可以找出一条合适的路径,从而提前抵达家门。

## 9.3　巴比涅原理与互补屏衍射

### 9.3.1　巴比涅原理

从菲涅耳衍射积分公式(9-1-8)可以导出一个有用的原理。考虑图9-3-1中的一对衍射屏(a)、(b),(a)屏的透过部分正是(b)屏的遮光部分,反之亦然,我们称它们为互补屏。

显然,$\Sigma_0=\Sigma_a+\Sigma_b$,故下式成立

$$\iint_{\Sigma_a}\mathrm{d}\Sigma+\iint_{\Sigma_b}\mathrm{d}\Sigma=\iint_{\Sigma_0}\mathrm{d}\Sigma \tag{9-3-1}$$

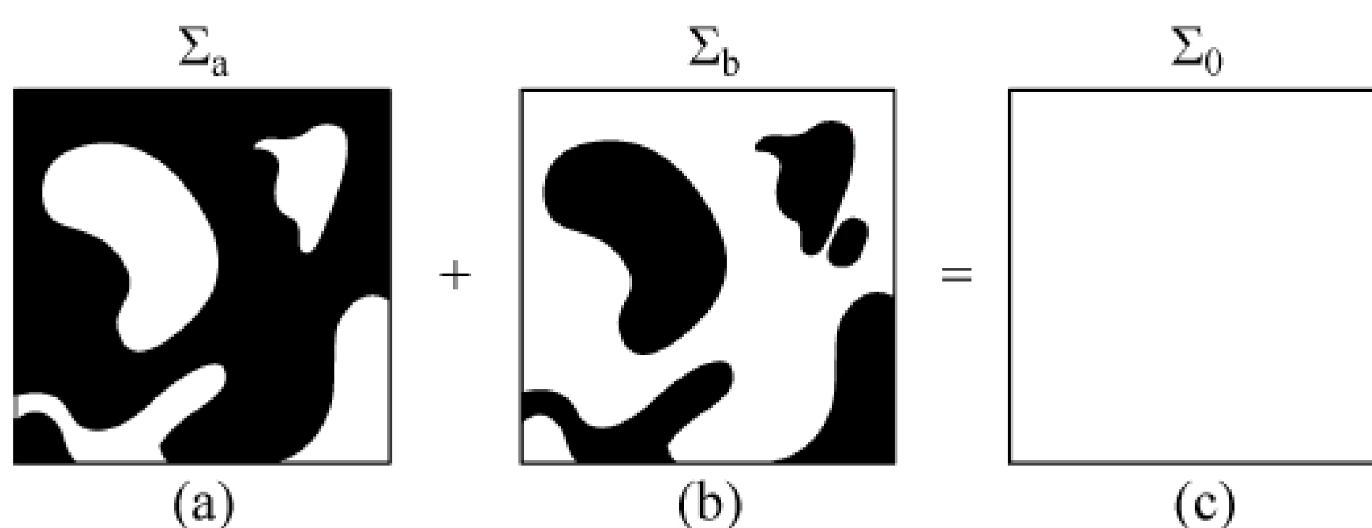


图9-3-1　巴比涅原理

上式左端第一项正是(a)屏的衍射场 $\tilde{E}_a(p)$,第二项是(b)屏的衍射场 $\tilde{E}_b(p)$,而右端是自由波场 $\tilde{E}_0(p)$(见图9-3-1(c)),于是

$$\tilde{E}_a(p)+\tilde{E}_b(p)=\tilde{E}_0(p) \tag{9-3-2}$$

它表明,互补屏造成的衍射场中复振幅之和等于自由波场的复振幅。这个结论称为巴比涅原理(Babinet's principle)。

### 9.3.2　巴比涅原理的应用

由于自由波场是容易计算的,因此利用巴比涅原理可以较方便地由一种衍射屏的衍射图样求出其互补屏的衍射图样来。

当 $\tilde{E}_a(p)=0$ 时,$\tilde{E}_b(P)=\tilde{E}_0(P)$,一个屏(a)的衍射强度为零点,在换上互补屏(b)的衍射强度与不存在屏(自由空间)的情况一样。

当 $\tilde{E}_0(P)=0$,即由点光源和两透镜(或两组透镜)构造的平行光衍射时,在接收屏上除了一个亮点(即点光源通过两个透镜后成的像点)外 $\tilde{E}_0(p)$ 皆等于零,从而除几何像点外,处处有

$$|\tilde{E}_a(p)|=|\tilde{E}_b(p)| \tag{9-3-3}$$

取它们各自复数共轭的乘积,得

$$I_a(p)=I_b(p) \tag{9-3-4}$$

亦即除光源点的几何像点之外,一对互补屏所形成的衍射图样的光强分布完全相同。

例如,一根细丝与一条狭缝的衍射花样,除零级中央主极大外,光强分布处处相同。这

就是动态丝径仪的原理(图 9-3-2)。

单缝衍射的第一极小 $\sin\theta=\dfrac{\lambda}{d}$,如图 9-3-2 所示,有几何关系 $\sin\theta\approx\tan\theta\approx\theta$ 和 $\tan\theta=\dfrac{W}{f}$,故所求细丝的直径为

$$d=\frac{f\lambda}{W} \tag{9-3-5}$$

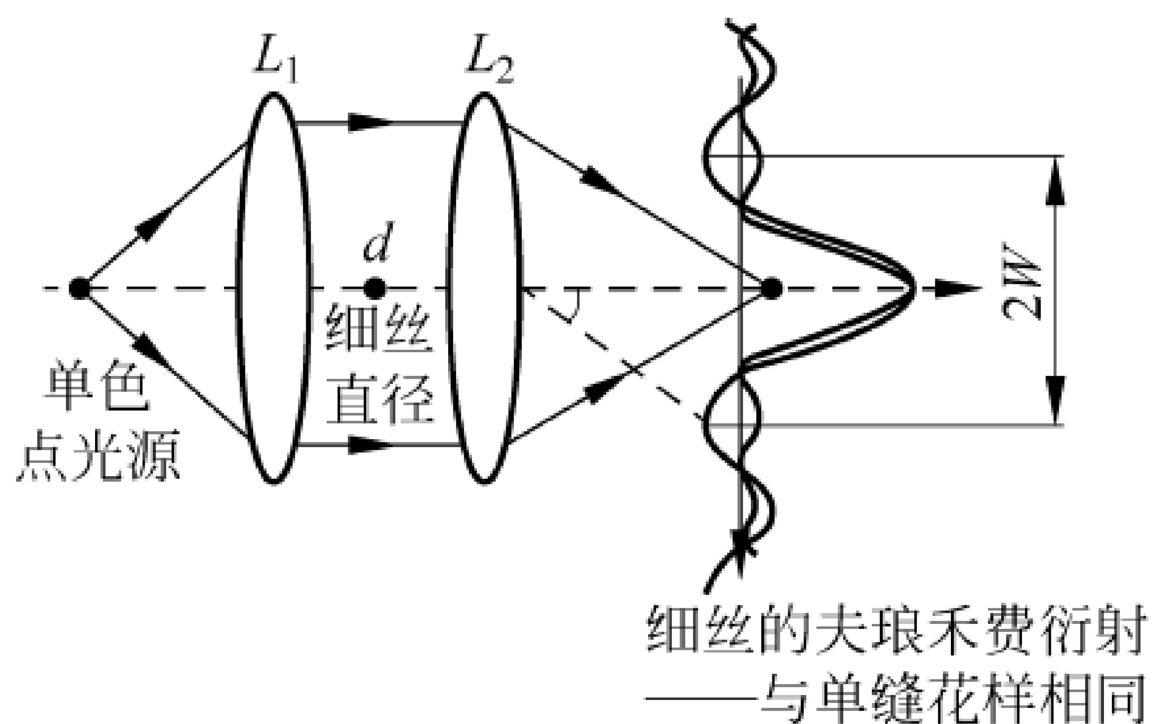


图 9-3-2 动态丝径仪的原理

于是,我们就可以通过对细丝直径衍射花样的监测及时掌握细丝制造的质量。这种激光动态丝径仪的精度比千分尺高一个数量级以上,而且测量时不需要接触细丝,不损害样品,便于做连续动态的检测,并可以使拉丝流程实现自动控制。

## 9.4 干涉与衍射的区别

光的干涉与光的衍射从本质上说是一样的,都是波的相干叠加的结果。所以,无论是应用叠加原理还是采用惠更斯-菲涅耳衍射积分公式,都遵循振动矢量叠加的原理。如果仅从这一点来看的话,实在没有必要将干涉和衍射加以区分。但实际上,光的干涉和衍射之间是既有区别又有联系的。光的干涉现象是两列或几列光波在空间相遇时,光强在一些区域加强,在另一些区域削弱,形成稳定的光强有规律分布的现象。光的衍射现象是光绕过障碍物偏离直线传播进入几何阴影,并在屏幕上出现光强不均匀分布的现象。下面我们以杨氏双缝干涉和夫琅禾费单缝衍射为例说明光的干涉和光的衍射之间的区别和联系。

### 9.4.1 杨氏双缝干涉

**1. 双缝干涉的原理**

使太阳光或某种单色光,通过一个具有单一狭缝的挡板,成为线光源,再让这一束光射到另一个具有相隔很近的两条狭缝的挡板上,而单缝到两条狭缝的距离相等,通过双缝的两束光就成了相干光源(即频率相同的两束光),当它们在屏幕上相遇时,相互叠加就形成了稳定的干涉条纹。

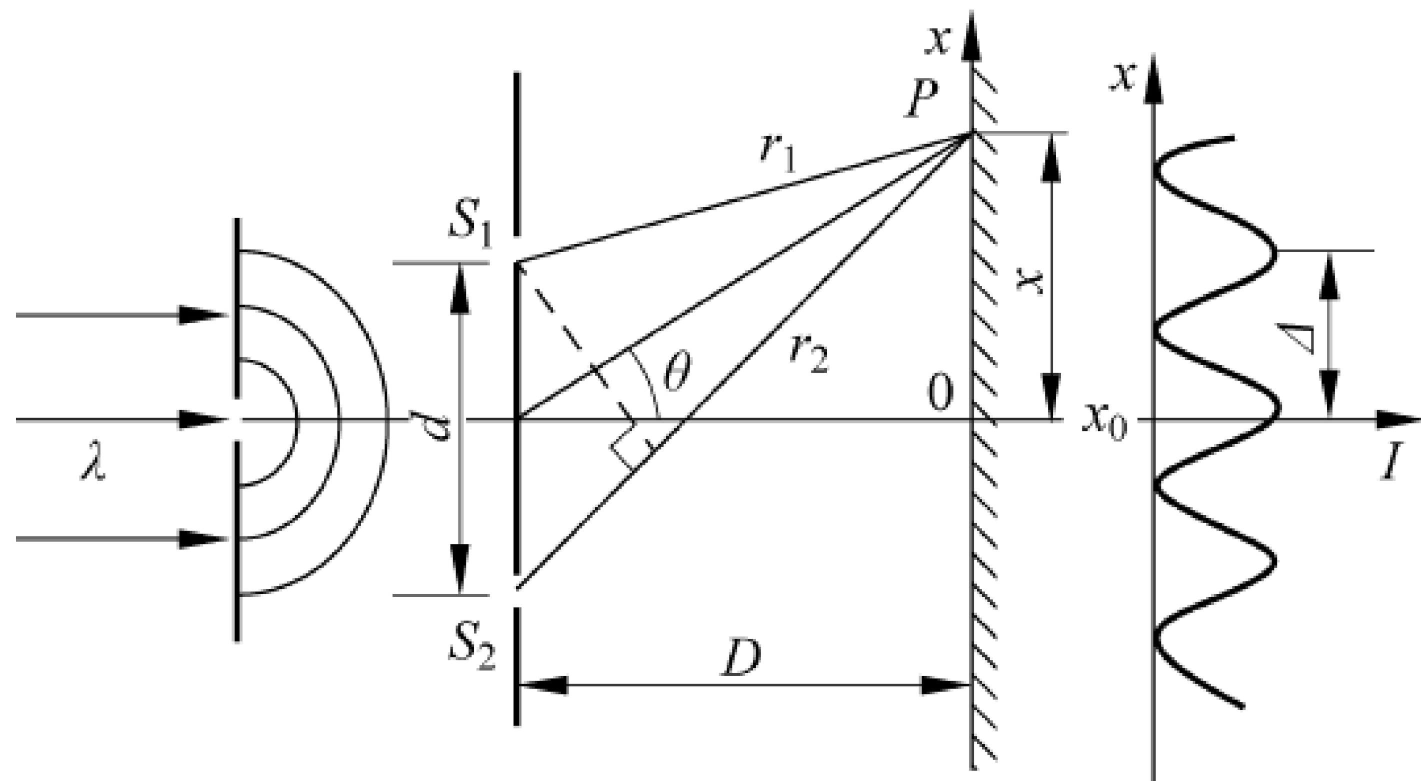


图 9-4-1 杨氏双缝干涉

**2. 双缝干涉的规律**

杨氏双缝干涉实验的实验装置如图 9-4-1 所示,用 $S_1$、$S_2$ 表示双缝,它们之间的距离为 $d$,双缝到屏幕之间的距离为 $D$,屏幕上某一点 $P$ 到双缝的距离分别为 $r_1$、$r_2$,则通过双缝 $S_1$、$S_2$ 的光到达屏幕上 $P$ 点的光程差 $\delta$ 就是

$$\delta=r_2-r_1 \tag{9-4-1}$$

当光的光程差 $\delta$ 是波长 $\lambda$ 的整数倍,即 $\delta=n\lambda(n=1,2,3,\cdots)$时,在 $P$ 点处将出现亮条纹;

当光的路程差 $\delta$ 是半波长 $\dfrac{\lambda}{2}$ 的奇数倍,即 $\delta=(2n+1)\dfrac{\lambda}{2}(n=1,2,3,\cdots)$时,在 $P$ 点处将出现暗条纹。

**3. 杨氏双缝干涉的光强分布**

狭缝 $S_1$ 和 $S_2$ 发出的光波单独到达屏上任一点 $B$ 处的振幅分别为 $A_1$ 和 $A_2$,光强分

别为 $I_1$ 和 $I_2$，则根据叠加原理，两光波叠加后的振幅和初相位分别为

$$A=\sqrt{A_1^2+A_2^2+2A_1A_2\cos(\varphi_2-\varphi_1)} \tag{9-4-2}$$

$$\tan\varphi=\frac{A_1\sin\varphi_1+A_2\sin\varphi_2}{A_1\cos\varphi_1+A_2\cos\varphi_2} \tag{9-4-3}$$

两光波叠加后的光强为

$$\bar{I}=\bar{A}^2=A_1^2+A_2^2+2A_1A_2\cos(\varphi_2-\varphi_1) \tag{9-4-4}$$

式中 $2A_1A_2\cos(\varphi_2-\varphi_1)$ 称为干涉项。

如果两振动相位相差是 π 的偶数倍，即 $\varphi_2-\varphi_1=2k\pi(k=0,1,2,\cdots)$，则 $\bar{I}=(A_1+A_2)^2$，合振动平均强度达到最大值，称为干涉相长；当振幅相同时，$I_{max}=4I_0$。

如果两振动相位相差是 π 的奇数倍，即 $\varphi_2-\varphi_1=(2k+1)\pi(j=0,1,2,\cdots)$，则 $\bar{I}=(A_1-A_2)^2$，强度达到最小值，称为干涉相消；当振幅相等时，$I_{min}=0$。

如果两振动的振幅相等且 $\varphi_2-\varphi_1$ 等于任何其他值，合振动的平均强度介乎上面两种情况之间，$\bar{I}=2A_1^2[1+\cos(\varphi_2-\varphi_1)]=4A_1^2\cos^2\dfrac{\varphi_2-\varphi_1}{2}$。所以，杨氏双缝干涉的光强分布图像如图 9-4-2 所示。

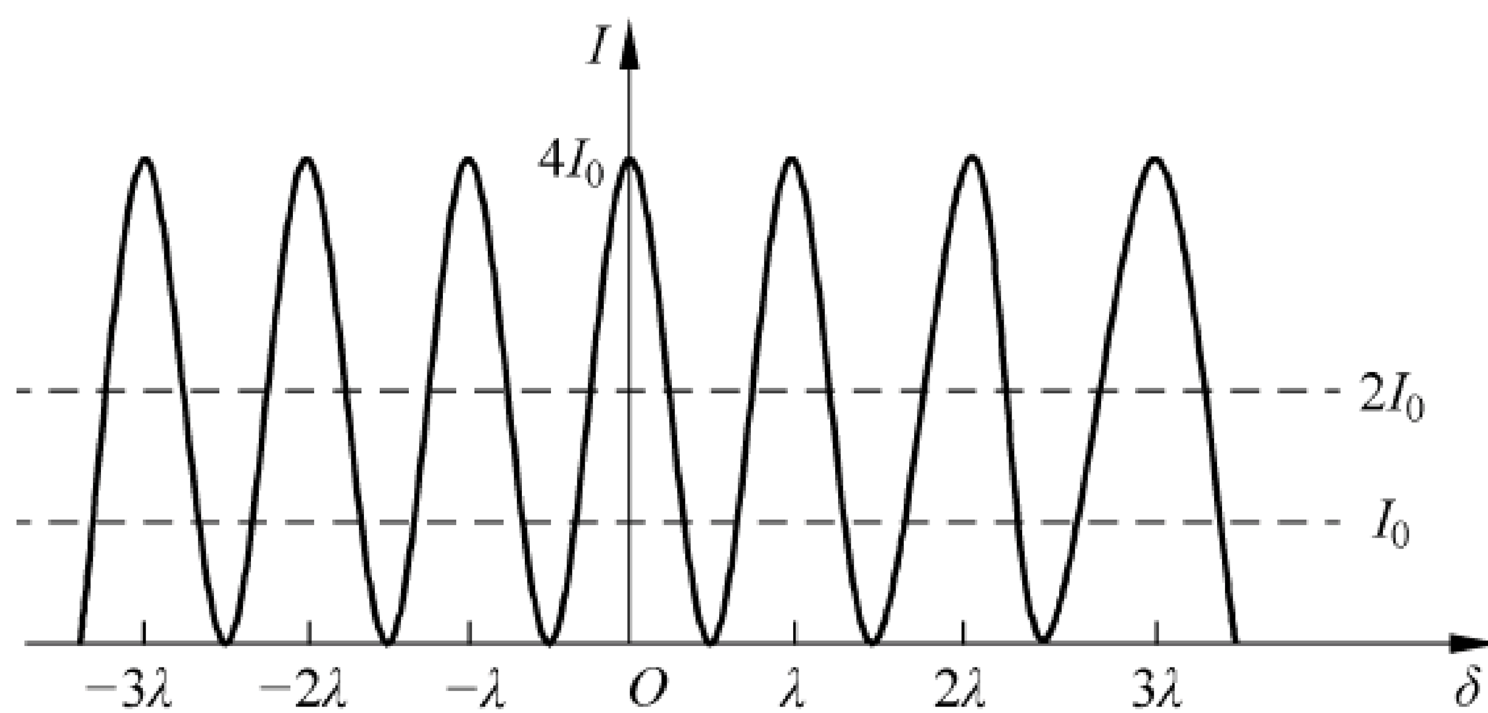



图 9-4-2　杨氏双缝干涉的光强分布

**4. 双缝干涉的现象**

利用白光做实验，在屏幕上得到的是彩色的干涉条纹；如果利用单色光做实验，则在屏幕上得到明暗相间的干涉条纹。其原因是：白光是由七种不同的单色光组成，每一种单色光所形成的条纹间距各不相同，红光最大，紫光最小，各种色光的干涉条纹叠加起来时，不能相互重叠，就得到了彩色的干涉条纹。

条纹间距离 $\Delta x$ 和双缝间距 $d$、双缝到屏幕的间距 $l$ 以及光的波长 $\lambda$ 之间的关系为

$$\Delta x=\frac{1}{d}\lambda \tag{9-4-5}$$

图 9-4-3 和图 9-4-4 所示分别为红光和白光的干涉条纹。

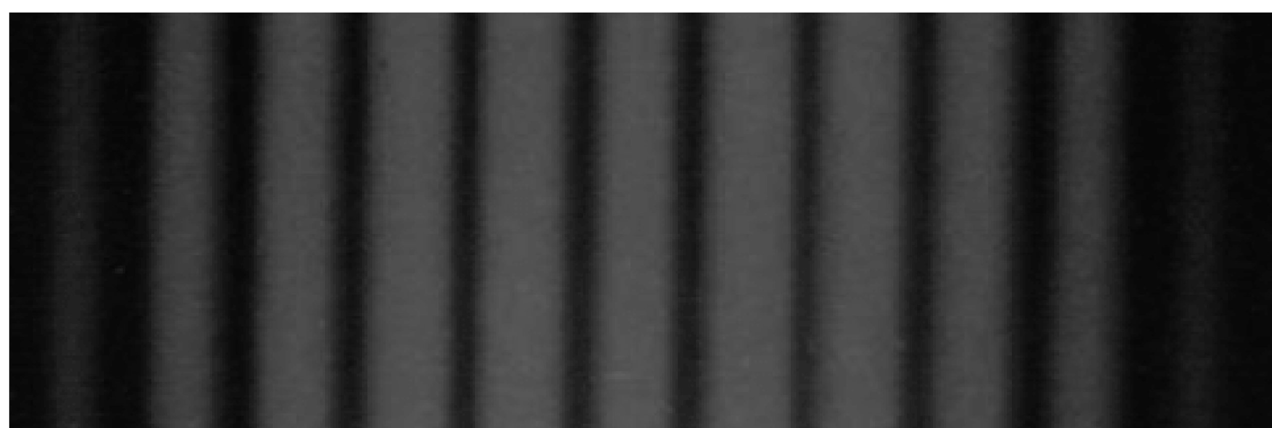


图 9-4-3　红光的干涉条纹

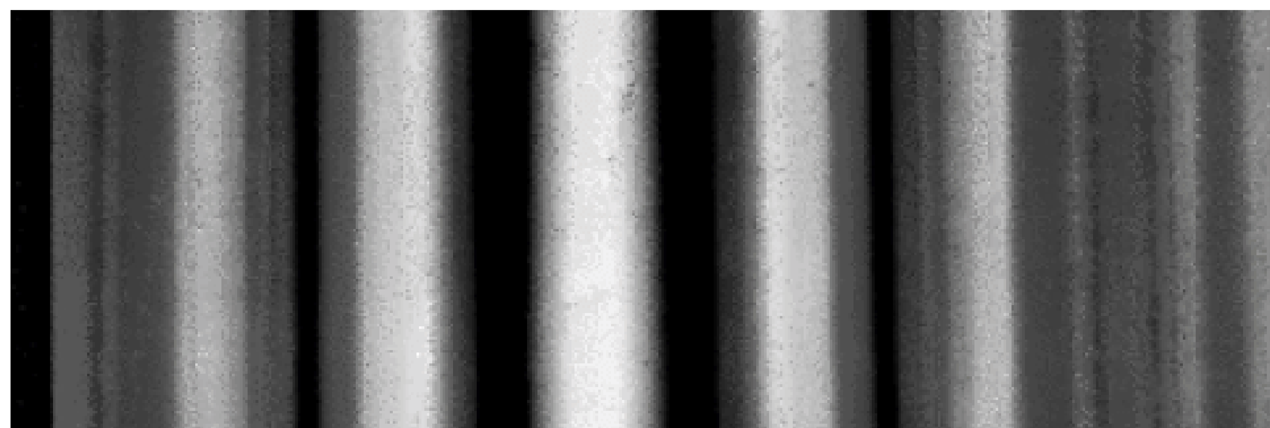


图 9-4-4　白光的干涉条纹

## 9.4.2 夫琅禾费单缝衍射

光的衍射现象是指光绕过障碍物偏离直线传播进入几何阴影,并在屏幕上出现光强不均匀分布的现象,如图 9-4-5 所示。

**1. 夫琅禾费单缝衍射的原理**

现在我们以夫琅禾费单缝衍射为例,说明光的衍射。夫琅禾费衍射指光源和观察屏离障碍物的距离均为无限远时所发生的衍射现象。夫琅禾费衍射又称平行光衍射,在实验中往往借助两个透镜来实现。图 9-4-6 是观察夫琅禾费单缝衍射的实验装置。

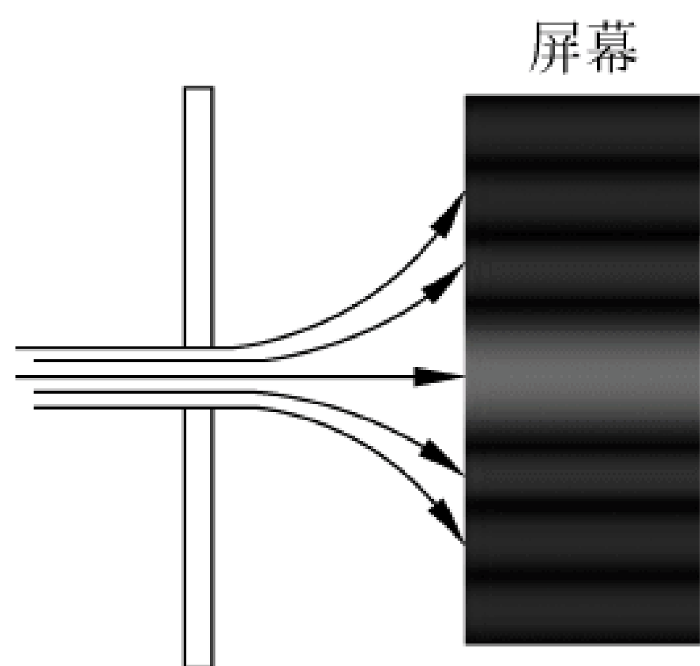



图 9-4-5 光的衍射

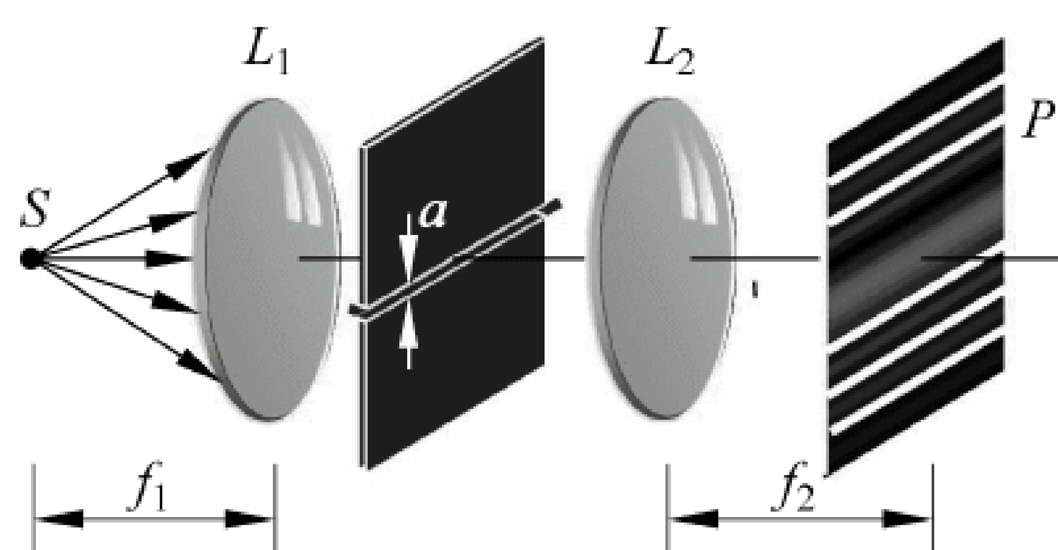



图 9-4-6 夫琅禾费单缝衍射的实验装置

在透镜 $L_1$ 的前焦点上放置单色点光源,点光源发出的光经过透镜折射后变成平行光,再垂直入射到开有一条狭缝的衍射屏上,缝的宽度为 $a$($a$ 略大于 $\lambda$),狭缝长度远大于宽度 $a$。入射光在狭缝发生衍射,透镜 $L_2$ 把无限远的衍射图样聚焦在它的后焦面上。从放在后焦面上的观察屏上能观察到夫琅禾费单缝衍射的衍射图样。由于狭缝的长度远远大于波长而缝宽很小,所以只在垂直狭缝的方向发生明显的衍射现象,即观察屏上的衍射图样只在 $x$ 方向上展开(图 9-4-7)。如果用与狭缝平行的线光源代替点光源,则在屏上会看到一组平行狭缝的明暗衍射条纹。



图 9-4-7 夫琅禾费单缝衍射图样

**2. 在夫琅禾费单缝衍射中,光强的分布**

(1) 单缝衍射中央最大值的位置

图 9-4-8 为夫琅禾费单缝衍射的实验原理图。单缝衍射中央最大值位置即零极衍射斑。光强的主极强位置出现在 $\theta=0$ 的地方,$\theta=0$ 相当于各衍射光线之间无光程差,根据费马原理,这就是几何光学像点的位置。由此我们可以看到"物像等光程差"的物理意义。几何光学中认为点光源发出的光线通过透镜后仍交于一点,这一点即点光源的像点。但从波动光学的角度看,光线会聚于一点,不一定产生很大的光强,还需要看各光线的相位关系。正是等光程性保证了到达像点各光线有相同的相位,从而产生最大的强度。即 $\sin\theta_0=0$(中央最大值的位置)就是在焦点 $P_0$ 处,$I_{P_0}=A_0^2$,光强为最大,这里各个次波相位差为零,所以振幅叠加相互加强。费马原理中所谓的"实际光线"就是零级衍射光线,而零级衍射斑的中心就是几何光学的像点。这是具有普遍意义的结论,利用它我们可以较容易地找

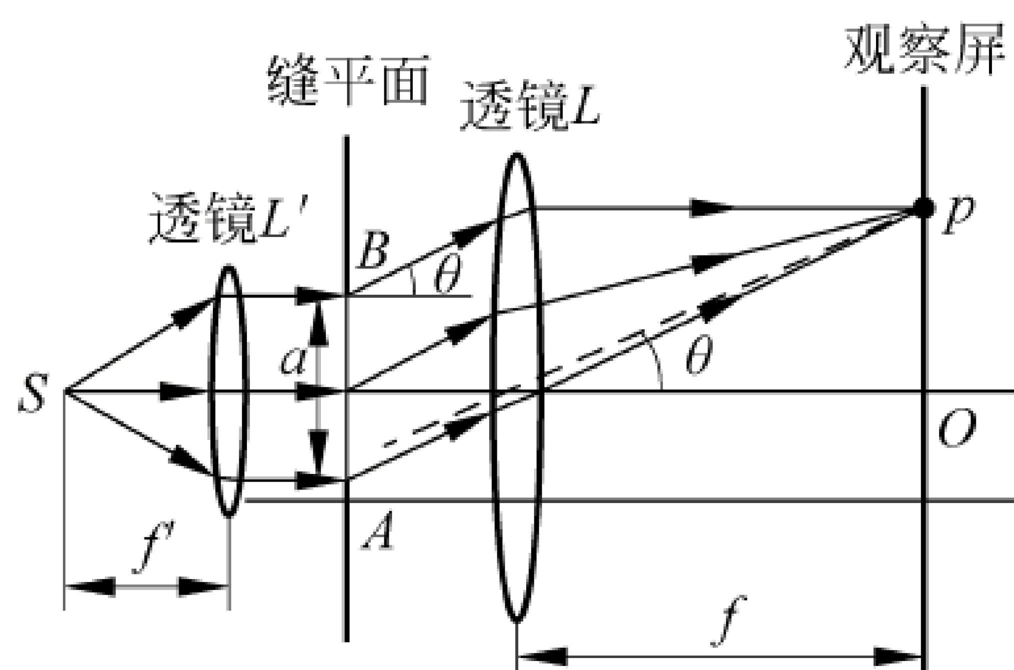



图 9-4-8 夫琅禾费单缝衍射的实验原理

到零级衍射斑的位置。

（2）单缝衍射最小值的位置

由 $\sin u=0$，解得满足 $u_k=\dfrac{\pi(a\sin\theta_k)}{\lambda}=k\pi$ 的一些衍射方向，即 $\sin\theta_k=k\dfrac{\lambda}{a}(k=\pm1,\pm2,\pm3,\cdots)$（最小值位置）时，$A_P$ 为零，屏上这些点是暗的。

（3）单缝衍射次极大的位置

在每两个相邻最小值之间有一最大值，即高级衍射斑。这些光强的次极大位置可由超越方程 $u=\tan u$ 解得。

由此可知分列于中央主最大两边的其他最大值（次极大）的位置为

$$\sin\theta_{10}=\pm1.43\frac{\lambda}{b}\approx\pm\frac{3}{2}\frac{\lambda}{b}$$

$$\sin\theta_{20}=\pm2.46\frac{\lambda}{b}\approx\pm\frac{5}{2}\frac{\lambda}{b}$$

$$\sin\theta_{30}=\pm3.47\frac{\lambda}{b}\approx\pm\frac{7}{2}\frac{\lambda}{b}$$

$$\vdots$$

$$\sin\theta_{k0}\approx\pm\left(k_0+\frac{1}{2}\right)\frac{\lambda}{b}\quad(k_0=1,2,\cdots)\tag{9-4-6}$$

综上所述，夫琅禾费单缝衍射光强分布曲线如图 9-4-9 所示。

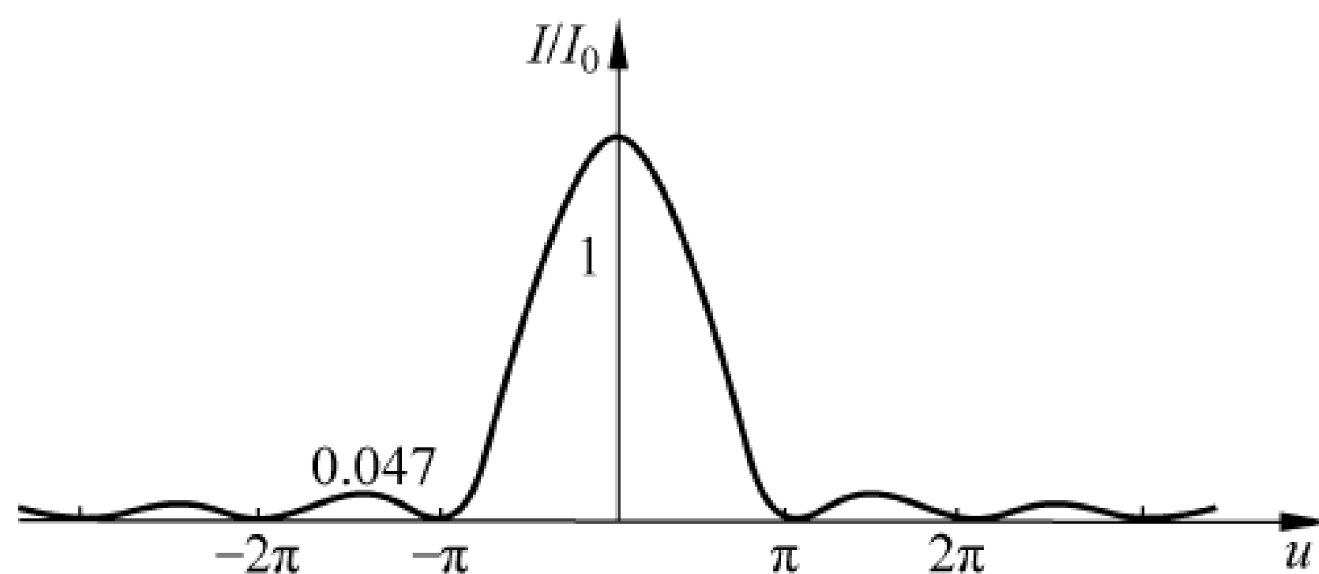



图 9-4-9　夫琅禾费单缝衍射的光强分布

### 9.4.3　干涉和衍射的区别

在上面的杨氏双缝干涉实验中，若把双缝之一遮蔽，那么光透过另一缝会在屏幕上出现什么现象呢？当然是单缝衍射条纹——条纹的亮度、宽度和分布都不均匀。除去上述的遮蔽，自然屏幕上会再次出现条纹的亮度、宽度和分布都基本均匀的干涉条纹。自然而然，我们会提出这样的问题：难道光（或者说光子）真的“知道”另一条缝是否被遮蔽吗？其实不然，根本的原因在于干涉和衍射现象的产生机理是相同的，都是子波源的相干叠加，都是光的波动性的明证。不同的是，干涉是有限个次波的相干叠加现象，而衍射则是无限个次波的相干叠加现象。除此以外，干涉和衍射现象还有很多明显的区别，我们通过上面所述杨氏双缝干涉实验和夫琅禾费单缝衍射实验作比较，可以得出以下几点：

（1）产生的条件不同

光的干涉现象的产生，必须满足相干光的条件，即频率相同，相位差恒定。例如从楔形肥皂膜上观察到的钠黄光的明暗相间条纹，或从水面油膜上观察到的彩色条纹，就属这类情况，从薄膜的前后表面反射出来的光就是相干光。

产生光的衍射需要满足的条件是：障碍物或孔的尺寸比光的波长小或跟光的波长相差不多。例如从小孔观察点光源或从狭缝观察线光源就属这种情况。光经过小孔或狭缝产生非直线传播的现象，此时便可在光屏上形成明暗相间的条纹。

（2）实验装置和方法不同

干涉是“光束”之间的相干叠加，要求每一条“光束”都要足够细，因而，各种干涉实验装

置中,每一个狭缝或针孔都要足够小,以至于可以认为其中仅包含有一个振动源。这些光束是有限条,或虽然有无限多条,但是光束之间是离散的、不连续的、可数的。

而衍射则是连续分布的无限多个点光源(次波中心)发出的光波的相干叠加。所以每一个衍射单元可以比较大,例如菲涅耳圆孔、夫琅禾费单缝、圆孔、矩孔等,其中每一个衍射单元中包含有许多(无限多)个扰动源。

(3) 采用的数学手段不同

由于干涉的“光束”是离散的,因而其叠加的过程在数学上既可以用求和的方法解决,又可以直接应用波的叠加原理;而衍射的扰动源是连续分布的,则在数学上必须采用积分(曲面积分)的方法求解,所以需要求解菲涅耳-基尔霍夫衍射积分公式。

(4) 图样产生机理不同

杨氏双缝干涉是双缝处发出的两列光波在屏幕上的叠加,当两列波到达屏幕上的某点的距离差等于波长的整数倍,该点是加强点,因而出现明条纹;当两列波到达屏幕上某点的距离差等于半波长的奇数倍,该点是减弱点,出现暗条纹。

夫琅禾费单缝衍射是光波从单缝处产生无数多个次波到达屏幕上的相互叠加,它们在屏幕上不同点叠加时,其相互减弱的程度有规律地变轻或变重是因为无数多个次波干涉的结果,明条纹、暗条纹强度以及出现的位置是无数多个次波干涉和单缝衍射调制后的共同效果。

(5) 物理结果不同

从最后的效果来看,干涉之后光的能量在空间均匀分布,无论杨氏干涉还是薄膜干涉,各级亮条纹有相差不大的空间角宽度和强度。以单色光为例,干涉图样互相平行且条纹宽度相同,中央和两侧的条纹没有区别。

而衍射所产生的一系列亮条纹中,有一个条纹具有较大的强度和角宽度。例如夫琅禾费单缝衍射条纹是平行不等距的,中央明条纹又宽又亮,两边条纹宽度变窄,亮度也明显减弱。又如,夫琅禾费圆孔衍射的爱里斑、单缝衍射的零级条纹,这个特殊的衍射级就在系统的几何像点上,因而衍射的结果更接近于几何成像的情况,或者说,几何光学就是衍射的零级近似。

## 9.5 衍射与几何光学仪器的分辨本领

因为几乎所有光学仪器的光学孔径都是圆形的,几何光学仪器的成像实际就是夫琅禾费圆孔衍射光强分布的零级最大,结合两物点发光的衍射零级最大角距离 $\delta\theta$ 与瑞利(Rayleigh John William Strutt,1842—1919)判据是分析几乎所有几何光学仪器分辨本领的基础。研究夫琅禾费圆孔衍射具有重要的实际意义。

### 9.5.1 圆孔衍射

光学仪器的光瞳通常是圆形的,所以讨论圆孔衍射对于分析光学仪器的衍射现象有特别重要的意义。

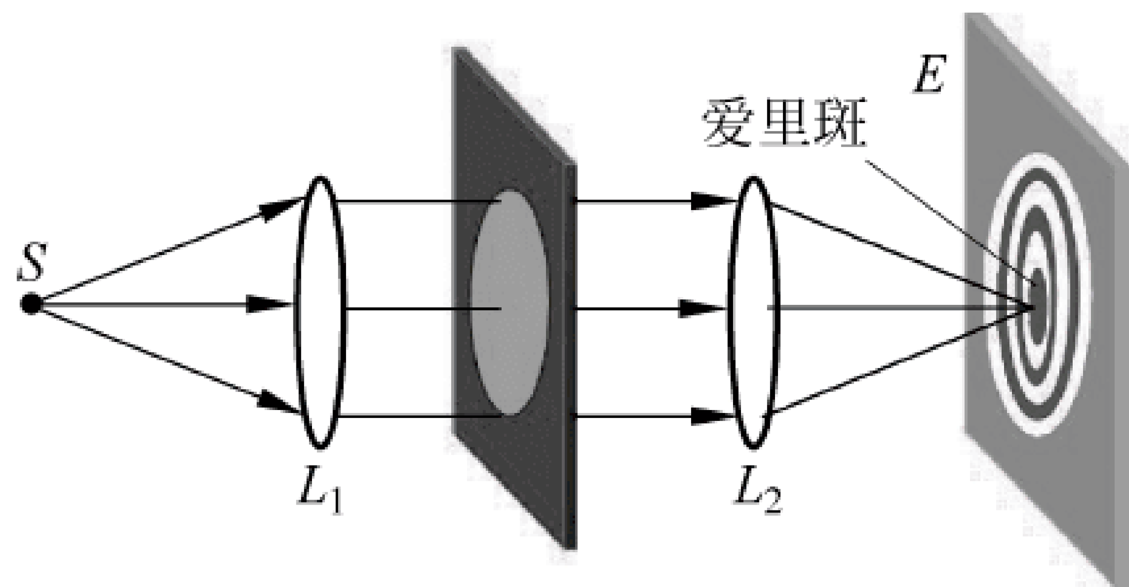


图 9-5-1 夫琅禾费圆孔衍射

如果把图 9-4-6 中的单缝换成半径为 $r$ 的圆孔,如图 9-5-1 所示。点光源 $S$ 放在准直透镜 $L_1$ 的物方焦点处,在透镜 $L_2$ 像方焦面处放置观察屏,则

在屏上可以得到夫琅禾费圆孔衍射图样。

按几何光学的规律，通过圆孔的光束经物镜 $L_2$ 聚焦后在观察屏上应该是一个亮点，由于圆孔衍射作用，中央有一亮斑，称为爱里斑，周围有越来越暗的亮环。约 84%的光能集中在中央亮斑上，第一亮环的光能占 7.2%，第二亮环的光能占 2.8%，第三亮环占 1.5%。

研究圆孔衍射的方法和单缝类似。由于圆孔的轴对称性，不难想象其衍射图样应该是与光轴对称分布的亮暗交替的圆环，计算时只要考虑衍射光强的径向分布就可以了。由于计算过程的数学比较复杂，这里只写出它的结果。图 9-5-2 是该计算结果的一部分。

$$\frac{I_\theta}{I_0}=\left[1-\frac{1}{2}\left(\frac{m}{1!}\right)^2+\frac{1}{3}\left(\frac{m^2}{2!}\right)^2-\frac{1}{4}\left(\frac{m^3}{3!}\right)^2+\frac{1}{5}\left(\frac{m^4}{4!}\right)^2-\cdots\right]^2 \tag{9-5-1}$$

其中 $m=\dfrac{\pi r\sin\theta}{\lambda}$。表 9-5-1 列出了圆孔衍射光强的极大、极小位置及其光强的相对值。

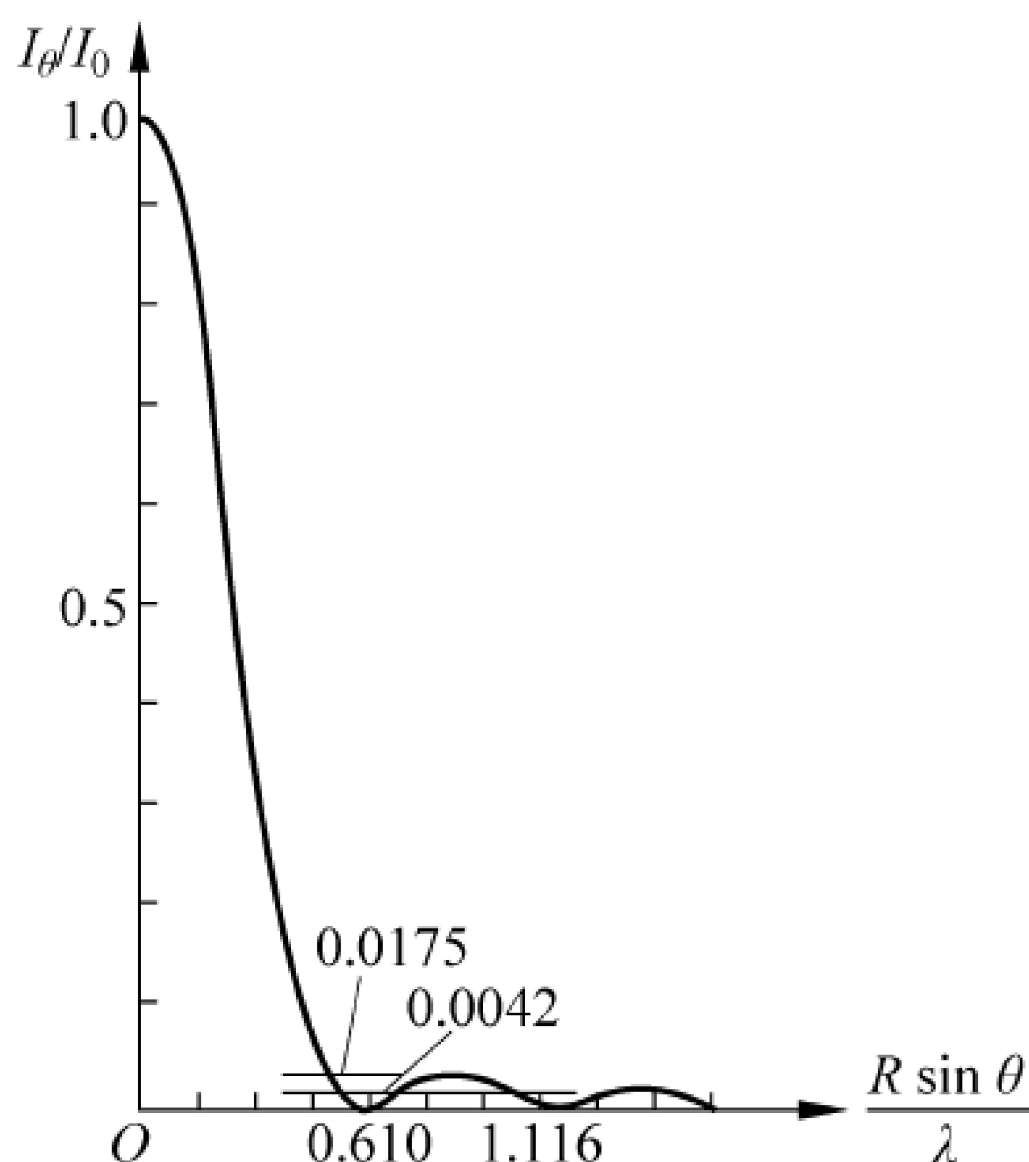



图 9-5-2　夫琅禾费圆孔衍射的光强分布

**表 9-5-1　夫琅禾费圆孔衍射的光强分布**

| | 中央主极大 | 第一极小 | 第一极大 | 第二极小 | 第二极大 | 第三极小 | 第三极大 |
|---|---|---|---|---|---|---|---|
| $\sin\theta$ | 0 | $0.61\dfrac{\lambda}{r}$ | $0.81\dfrac{\lambda}{r}$ | $1.12\dfrac{\lambda}{r}$ | $1.33\dfrac{\lambda}{r}$ | $1.62\dfrac{\lambda}{r}$ | $1.85\dfrac{\lambda}{r}$ |
| $\dfrac{I_\theta}{I_0}$ | 1 | 0 | 0.0175 | 0 | 0.00415 | 0 | 0.0016 |

夫琅禾费圆孔衍射图样中央是亮斑，周围亮环的强度下降比单缝时要快，实际上只能看到一、二个亮环，从表 9-5-1 中数据可以得出中央亮斑(爱里斑)的角半径 $\Theta$ 为

$$\Theta=\frac{0.61}{r}\lambda=\frac{1.22}{d}\lambda \tag{9-5-2}$$

式中 $d$ 是圆孔直径，$r$ 是半径。

## 9.5.2　光学成像仪器的分辨本领

在几何光学里，物点经理想透镜能得到一个像点。由于存在衍射，几何光学中的所谓像点实际上是爱里斑(图 9-5-3)，这就限制了成像系统分辨两相近物点的能力。通常用分辨本领或最小可分辨物点间距来描述这种能力。

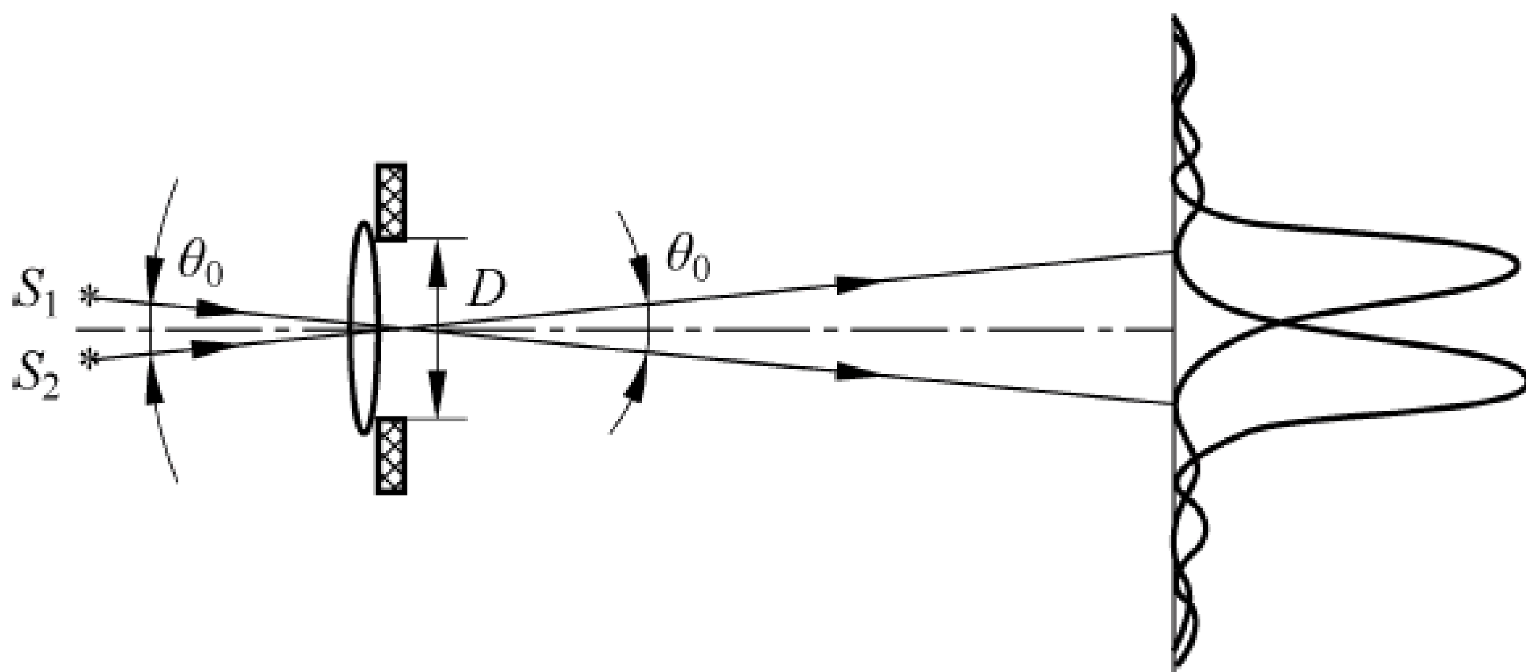



图 9-5-3　成像仪器的分辨本领

**1. 瑞利判据**

如果两物点是强度相等的同频点光源,当它们离得很开,经成像系统后两个爱里斑的角距离 $\theta_0$ 大于爱里斑角半径 $\Theta$ 时,毫无困难分辨,如图 9-5-4(a)所示;如果两物点很靠近,爱里斑重叠到 $\theta_0<\Theta$ 时,完全不能分辨,如图 9-5-4(c)所示,图中粗线是两个爱里斑非相干叠加时合光强分布,和单一物点的爱里斑没有区别。

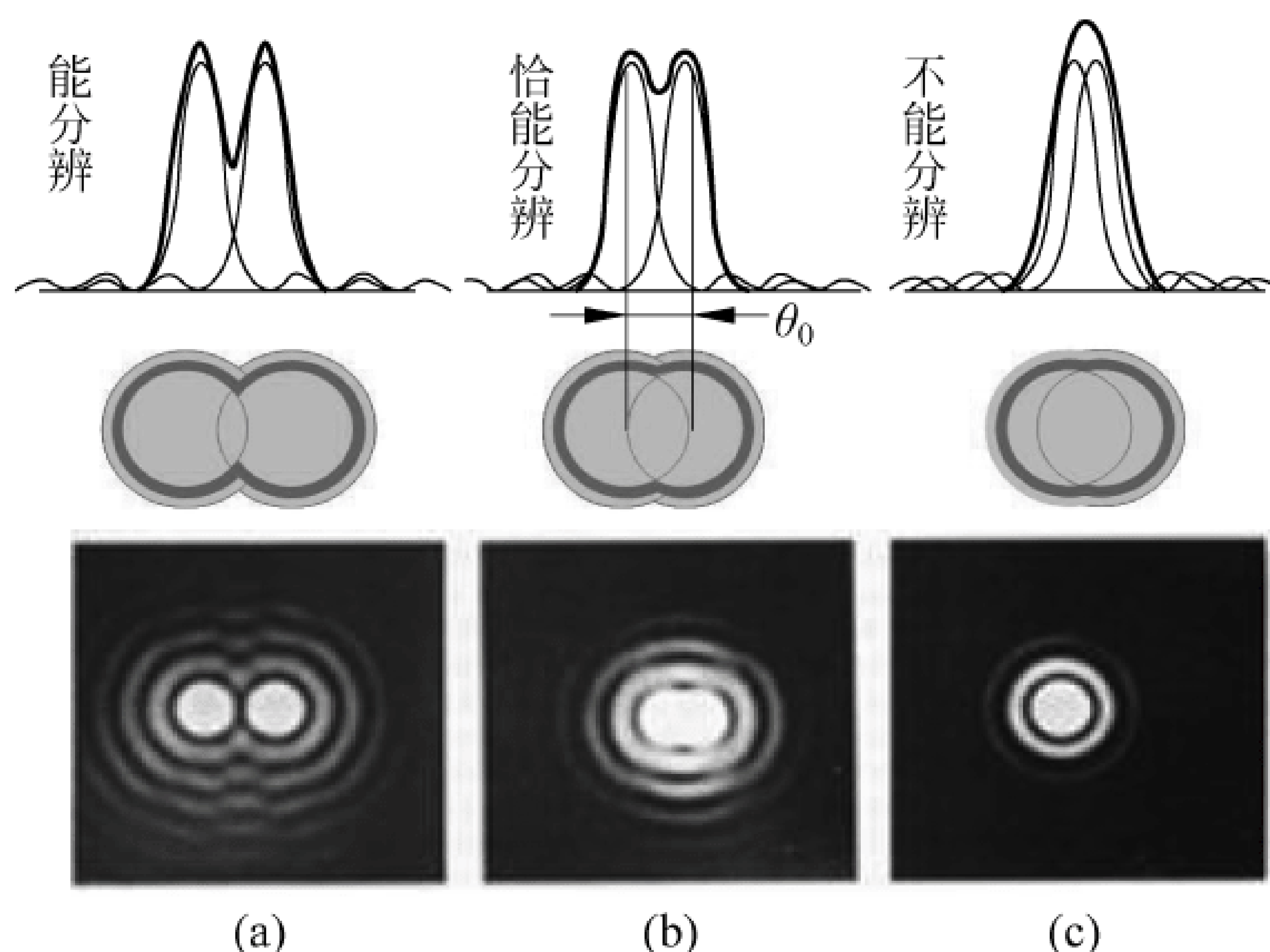



图 9-5-4 瑞利判据说明图

如果两物点靠近到两爱里斑角距离刚好等于爱里斑角半径 $\theta_0=\Theta$ 时,如图 9-5-4(b)所示,这时合光强分布曲线中凹陷处强度为两侧峰值的 74%,正常眼均能分辨出是两个部分重叠的爱里斑。瑞利提出:可作为成像系统对两物点恰可被分辨的准则,这个准则通常称为瑞利判据。

按瑞利判据,恰可被分辨两物点的角距离 $\delta\theta$,或称最小可分辨角为

$$\delta\theta=\frac{0.61}{r}\lambda=\frac{1.22}{d}\lambda \tag{9-5-3}$$

式中 $r$、$d$ 分别为仪器通过孔的半径和直径。从式(9-5-3)出发我们可以讨论人的眼睛、望远镜、显微镜、照相机等几种成像系统的分辨本领。

**2. 眼睛**

用眼睛观察远处物体时,视网膜上的像实际上是自物体发出的光通过眼睛瞳孔而产生的夫琅禾费圆孔衍射。用图 9-5-5 简化眼模型,满足瑞利判据要求的两物点 $A$、$B$ 对简化眼睫点 $C$ 的张角 $\delta\theta_e$ 由式(9-5-3)给出,此即人眼的最小可分辨角。若物体放在明视距离处,则最小可分辨物点间距 $\delta y_e$ 为

$$\delta y_e=l_{明}\,\delta\theta_e \tag{9-5-4}$$

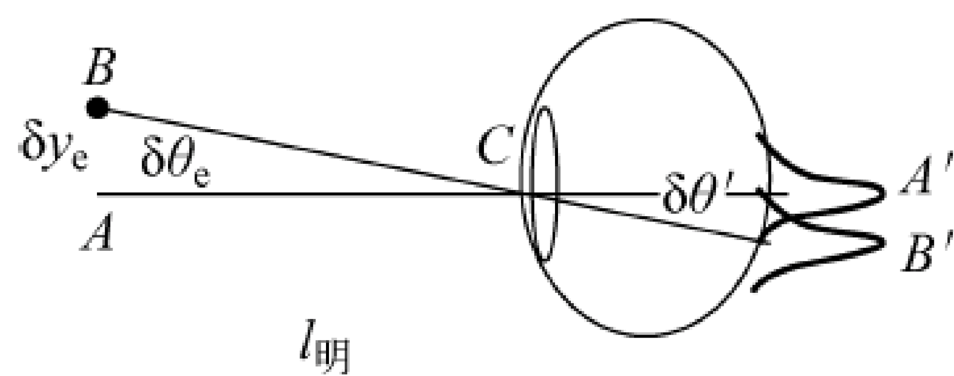



图 9-5-5 人眼折光系统的分辨本领

光照充足时(~100 勒克斯照度下),人眼瞳孔直径 $d$ 约为 2.3mm。取人眼最灵敏度波长 $\lambda=550\text{nm}$,由式(9-5-3)得人眼的最小可分辨角 $\delta\theta_e$ 和最小可分辨物点间距 $\delta y_e$ 分别为

$$\delta\theta_e=\frac{1.22}{d}\lambda=2.9\times10^{-4}\,\text{rad}\approx1' \tag{9-5-5}$$

$$\delta y_e=\delta\theta_e\times250\text{mm}=72.5\times10^{-3}\,\text{mm}\approx75\mu\text{m} \tag{9-5-6}$$

人眼的分辨能力由眼睛折光系统的分辨能力和视网膜的分辨能力两者决定。$\delta\theta_e$、$\delta y_e$只反映了前者，后者与视网膜上感光细胞的大小和间隔有关。为了能分辨是两个像点，至少要求它们不落在同一视神经细胞上。简化眼睫点 $C$ 到视网膜距离为 17.1mm，视角为 $2.9\times10^{-4}$rad 的两物点，落在视网膜上的间距 $\delta y'$为

$$\delta y' = (17.1\times2.9\times10^{-4})\text{mm} = 5\mu\text{m} \tag{9-5-7}$$

这恰好等于视神经细胞的间隔(视神经细胞直径为 1～3μm)。人眼最小分辨角和视神经细胞分布这么巧妙地匹配，显示人眼对自然规律适应之精细，结构之高超。

虽然上面算出人眼的最小可分辨角 $\delta\theta_e\approx1'$，最小可分辨物点间距 $\delta y_e\approx75\mu\text{m}$，但实际上人们常取 $\delta\theta_e=2'\sim4'$，$\delta y_e=150\sim300\mu\text{m}$。

**3. 望远镜**

望远镜是用来观察远处的物体的，物镜的边框为入瞳，起衍射孔的作用。物镜的最小可分辨角 $\delta\theta$ 由式(9-5-3)决定。设物镜入瞳直径为 $D$，则

$$\delta\theta = \frac{1.22\lambda}{D} \tag{9-5-8}$$

任何一个实际使用的望远镜，物镜的入瞳直径 $D$ 总是大大超过人眼瞳孔直径 $d$，所以，用望远镜观察远处物体时，会明显提高对物体的分辨本领，所提高的倍数是 $D/d$。物镜越大，则分辨本领提高的倍数越大。$D$ 为 6m 的天文望远镜物镜，最小分辨角约为百分之二秒。

望远镜的分辨本领主要由物镜的分辨本领决定，目镜的作用只是将物镜所成的初像再放大，以保证将物镜的最小可分辨角 $\delta\theta_o$ 放大到人眼最小可分辨角 $\delta\theta_e$ 的程度。换句话说，为充分发挥物镜的分辨能力，望远镜的视角放大率应等于 $\delta\theta_e/\delta\theta_o$，并称为正常(有效)放大率，记作 $M_n$，即

$$M_n = \frac{\delta\theta_e}{\delta\theta_o} \tag{9-5-9}$$

**4. 显微镜**

显微镜是用来观察放在物镜物方焦点邻近的细小物体的，因而物点发出的光束经物镜后几乎是一组平行光束，而位于物镜像方焦面处的孔阑(兼出瞳)起衍射孔的作用，见图 9-5-6。物镜像平面和出瞳距离(显微镜光学间隔 $\Delta$)比物镜出瞳直径 $P'_1P'_2=D'$要大得多，因此可以看成夫琅禾费圆孔衍射，只是这个圆孔要理解为物镜的出瞳，式(9-5-3)中 $d$ 要用出瞳直径 $D'$代替，即显微镜的最小可分辨像点角距离为

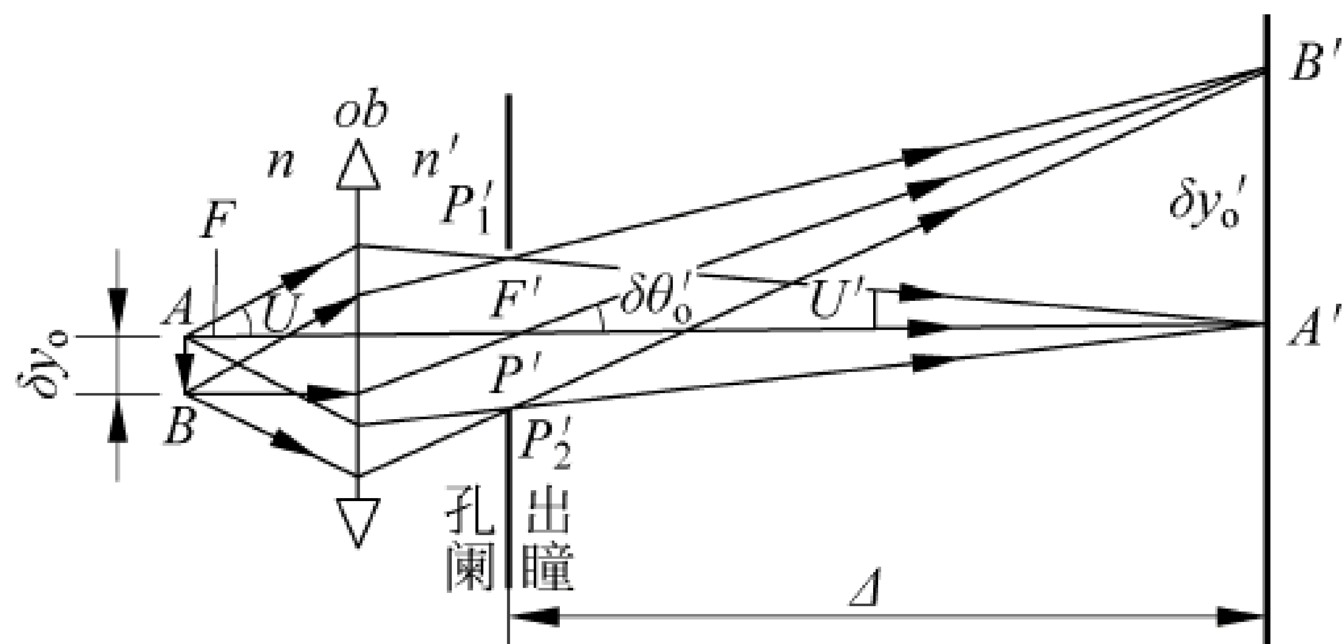


图 9-5-6　显微镜物镜的分辨本领

$$\delta\theta'_o = \frac{1.22\lambda}{D'} \tag{9-5-10}$$

显微镜的分辨本领常用物镜的最小可分辨物点间距 $\delta y_o$ 表示。图 9-5-6 中表示 $\delta y'_o$与 $\delta y_o$ 共轭的最小可分辨像点间距。

$$\delta y'_o = \delta\theta'_o\Delta = 1.22\lambda\frac{\Delta}{D'} = \frac{0.61\lambda}{\sin U'} \tag{9-5-11}$$

式(9-5-11)中用了 $\tan U'=\sin U'$近轴近似，校正了彗差的物镜基本上满足正弦条件

$\delta y n\sin U=\delta y'n'\sin U'$,而显微镜中 $n'$ 总是等于 1,于是,可得物镜最小可分辨物点间距为

$$\delta y_o=\frac{0.61\lambda}{\mathrm{N.A.}} \tag{9-5-12}$$

其中,$\mathrm{N.A.}=n\sin U$ 称为物镜的数值孔径。干燥系物镜 N. A. 在 1 以下,$\delta y_o\geqslant 0.61\lambda$;油浸物镜的 N. A. 可到 1.6,$\delta y_o$ 可到 $0.4\lambda$。

物镜的最小可分辨角 $\delta\theta_o=\dfrac{\delta y_o}{l_{明}}$。显微镜的有效放大率 $M_n$ 可按将视角$\dfrac{\delta y_o}{l_{明}}$恰好放大到 $\delta\theta_e$ 得出,即

$$M_n=\frac{\delta\theta_e l_{明}}{\delta y_o}=\frac{\delta y(\mathrm{N.A.})}{0.61\lambda} \tag{9-5-13a}$$

若选取 $\delta y_e=150\sim300\mu\mathrm{m}$,$\lambda=0.55\mu\mathrm{m}$,上式可以写为

$$M_n=447\mathrm{N.A.}\sim 895\mathrm{N.A.} \tag{9-5-13b}$$

为方便记忆,在实用上常近似地将上式写成

$$M_n=500\mathrm{N.A.}\sim 1000\mathrm{N.A.} \tag{9-5-13c}$$

500N. A. ～1000N. A. 的范围称为“有效放大倍数范围”。物镜一经选定,选择目镜放大倍数应该使得($\beta_{ob}\cdot M_{oc}$)在 500N. A. ～1000N. A. 之间。目镜如果选得不合适,($\beta_{ob}\cdot M_{oc}$)不足 500N. A. 表示未能充分发挥物镜的分辨本领;若($\beta_{ob}\cdot M_{oc}$)超过 1000N. A. ,称为“无效放大”,在此情况下不能看到在有效放大倍数内未能分辨的细节,由于像差变严重,不仅无益反而有害。显微镜的作用是分辨细节,所以不能单用放大率 $M_n$ 表征,若不考虑物镜数值孔径 N. A. 常会造成错误。

**5. 照相机**

除翻拍相机外,通常是物距较焦距大许多倍的情况下成像。普通相机物镜的焦距为 50～100mm,相机底片上的像是由物镜入瞳产生的夫琅禾费衍射图样。物镜的最小可分辨角 $\delta\theta_o$ 可用式(9-5-3)计算。相机分辨本领既取决于照相物镜的最小可分辨角,也取决于底片上感光乳剂的结构和性质。为充分利用照相物镜的分辨本领,要求感光乳剂对于相差不多的光强有足够灵敏的感光反应,即有较高的反衬灵敏度。此外,乳剂结构还应该足够细密,以致乳剂面上感光元的大小和相邻感光元间的距离都要小于物镜所产生的爱里斑的直径。当乳剂满足上述两方面要求时,照相机的分辨本领才可以说取决于物镜的分辨本领。

相对孔径为 $D/f'$ 的照相物镜,其最小分辨像点间距为

$$\delta y'_o=f'\delta\theta_o=1.22\lambda\frac{f'}{D} \tag{9-5-14}$$

式(9-5-14)中 $f'$ 为物镜像方焦距,$D$ 可以理解为物镜的直径,实际上是物镜出瞳直径或光圈的开孔大小。

为充分利用物镜的分辨本领,感光乳剂单位长度内能分辨的线条数为

$$\frac{1}{\delta y'_o}=\frac{1}{1.22\lambda}\cdot\frac{D}{f'} \tag{9-5-15}$$

式(9-5-15)表示,相对孔径越大的照相物镜,其分辨本领越大。为充分利用它的分辨本领,要求底片单位长度内能分辨的线条数也越多。例如:最大相对孔径为 1∶3.5 的照相物镜,若取 $\lambda=0.55\mu\mathrm{m}$,则要求用每毫米能分辨 420 条线的底片。一般的照相物镜剩余像差很大,它显著地降低了物镜的分辨本领,因而对感光底片的要求可能较 420 条/mm 小许多,一般要求 200 条/mm 就足够了。

# 9.6　三基色原理与彩色电视系统

颜色是不同波长的可见光辐射作用于人的视觉器官后所产生的心理感受。可见光的白光是由红、橙、黄、绿、青、蓝、紫七种光谱色组成的，而各种颜色又可以由这七种光谱色按不同的比例构成。人脑有记忆、联想等功能，因此人观察到的颜色，往往受有关颜色的经验、背景颜色及物体形状等心理因素的影响。所以，颜色是一种和物理、生理及心理学有关的复杂现象。

## 9.6.1　光的色散

大雨过后，我们经常为迷人的七色彩虹而欢呼，但又有多少人能想到在七彩的背后，究竟隐藏了多少奥秘？我们知道，太阳光是由多种不同频率的色光所组成的。在均匀介质中，它们沿着相同的方向传播，因此，我们平时所看到的太阳光就是白光。但下雨后，空气中混有水分，空气的成分就相差很远，不再是近似的均匀介质，因此太阳光在传播过程中就发生折射，各种光发生折射。由于各种光的频率不同，造成它们的传播速度不同，因此，它们在有水分的空气中所走过的路程不同，不同频率的光将偏离原来的传播方向，这就形成我们在下雨后所见到的彩虹。这也是光的色散的体现。所谓光的色散，就是复色光分解成单色光而形成光谱的现象。白光通过三棱镜后会折射为七种颜色，就是光的色散的现象，如图 9-6-1 所示。

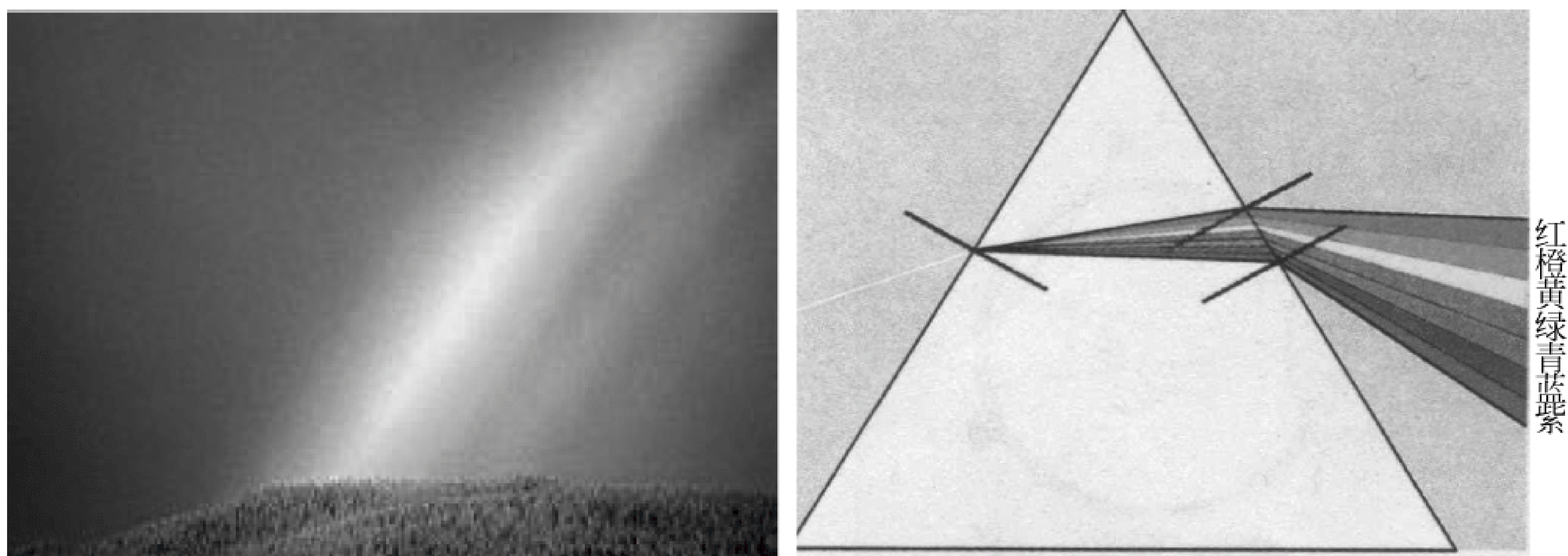



图 9-6-1　光的色散

## 9.6.2　三基色原理

### 1. 色光三基色(三原色)与三补色

颜色是光在人眼中的反映。日常生活中，我们所遇到的颜色多种多样：红色、粉红色、橙色、浅橙色、黄色、淡黄色、绿色、淡绿色、蓝色、深蓝色、紫色等。这众多的颜色有什么规律吗？通过实验，人们发现：红＋绿＝黄，红＋蓝＝紫，蓝＋绿＝青，红＋蓝＋绿＝白……，其中，红、绿、蓝三种颜色合成的颜色最多。于是人们得出以下结论：所有的颜色都是由红、绿、蓝三种基本颜色合成的，称红、绿、蓝为三基色或三原色。这三种颜色是相互独立的，其中任何一种颜色均不能由其他二种颜色混合产生。它们又是完备的，即所有其他颜色都可以由三基色按不同的比例组合而得到。

能够匹配所有颜色的三种颜色，称为三原色。其颜色为：红(R)、绿(G)、蓝(B)三种。匹配实验表明，能够匹配所有颜色的三种颜色不是唯一的。人们通常选用红(R)、绿(G)、蓝(B)作为三原色，其原因可能是：用不同量的红、绿、蓝三种颜色直接混合，几乎可以得到经

常使用的所有颜色；红、绿、蓝三种颜色恰与人的视网膜上红视锥、绿视锥和蓝视锥细胞所敏感的颜色相一致。

对于白光而言，与三原色互补的颜色称为三补色。三补色也能够匹配所有颜色。绿的补色是品红，蓝的补色是黄，红的补色是青。其匹配原理如图 9-6-2 所示。

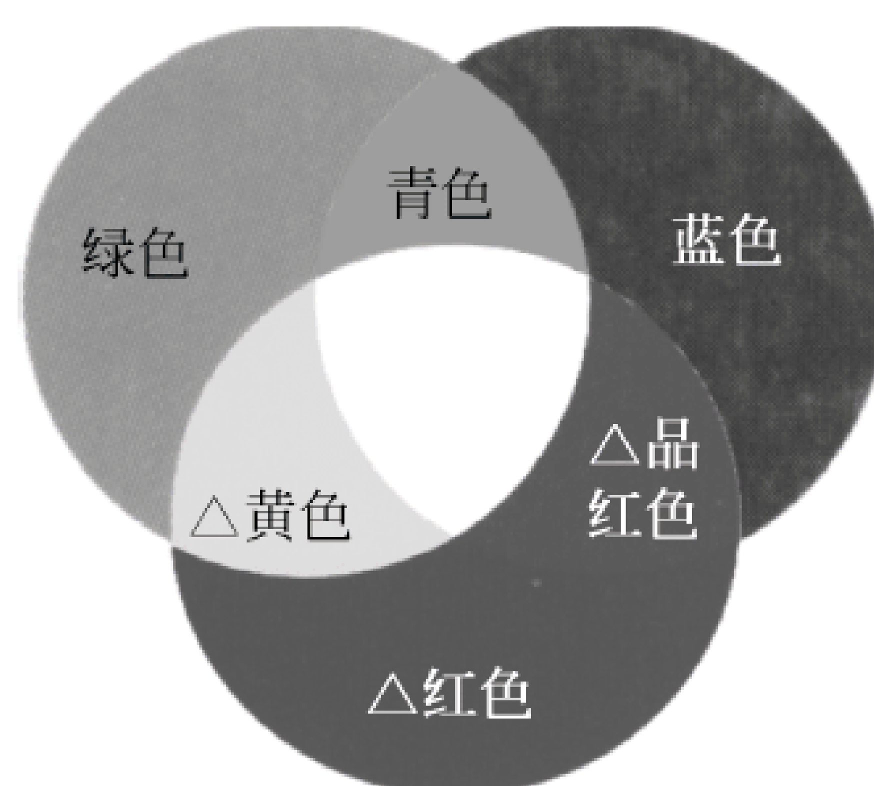



图 9-6-2 三原色与三补色的匹配原理

**2. 三基色原理**

通过混色现象和配色实验，人们总结出了三基色原理：可见光谱中的大多数颜色感觉，可通过红色(R)、绿色(G)和蓝色(B) 3 种色光按不同强度比例混合而获得。即任意彩色光 $F$，其配色方程可写成：$F=R[R]+G[G]+B[B]$，其关系由配色实验实际测量的配色数据确定。其中 $R$、$G$、$B$ 为三基色的刺激值，[R]、[G]、[B]为三基色的原色刺激。根据色度学的标准规定，[R]代表波长为 700nm，光通量为 1W 的红光作为 1 个红基色单位；[G]代表波长为 546.1nm，光通量为 4.5907W 的绿光作为 1 个绿基色单位；[B]代表波长为 435.8nm，光通量为 0.0601W 的蓝光作为 1 个蓝基色单位。

三基色的具体内涵为：第一、三基色按一定比例混合可以得到大自然的全部彩色，同样大自然的全部彩色都可以分解为三基色。第二、三基色必须是独立的一种颜色，任何其他颜色均不可能混合出三基色其中的一种颜色。

三基色原理奠定了彩色显像的基础理论和技术依据。人眼对红、绿、蓝最为敏感，人的眼睛就像一个三色接收器的体系。故光电显示彩色技术采用了红绿蓝作为三基色，不论 CRT 方式或 LCD 方式以及 DLP、LCOS 等，均采用了 RGB 三基色来合成彩色图像，只不过合成的方式有所不同，三枪 CRT 及三片 LCD 技术是采用了直接混色法，单枪 CRT 及单片 LCD 采用的是空间混色法，DLP、LCOS 技术则采用了时间混色法。

**3. 颜料三基色**

讲到绘画、图像，自然离不开颜色，所有的图案都是由基本形状和颜色组成，颜色构成了我们图像处理的一个重要部分。

(1) 颜料原理——颜色的混合与组成

一些常见的颜色的组成：

红色＋绿色＝黄色
绿色＋蓝色＝青色
红色＋蓝色＝品红
红色＋绿色＋蓝色＝白色

黄色、青色、品红都是由两种基色相混合而成，所以它们又称相加二次色。

(2) 颜料三基色(三补色)

通过实验发现，颜色的混合与组成有如下规律：

红色＋青色＝白色
绿色＋品红＝白色
蓝色＋黄色＝白色

所以，青色、黄色、品红分别又是红色、蓝色、绿色的补色。由于每个人的眼睛对于相同的单色的感受有不同，所以，如果我们用相同强度的三基色混合时，假设得到白光的强度为

100%，这时候人的主观感受是：绿光的强度最大，红光次之，蓝光最弱。

除了相加混色法之外还有相减混色法。在白光照射下，青色颜料能吸收红色而反射青色，黄色颜料吸收蓝色而反射黄色，品红颜料吸收绿色而反射品红。也就是：

白色－红色＝青色

白色－绿色＝品红

白色－蓝色＝黄色

另外，如果把青色和黄色两种颜料混合，在白光照射下，由于颜料吸收了红色和蓝色，而反射了绿色，对于颜料的混合我们表示如下：

颜料(黄色＋青色)＝白色－红色－蓝色＝绿色

颜料(品红＋青色)＝白色－红色－绿色＝蓝色

颜料(黄色＋品红)＝白色－绿色－蓝色＝红色

以上的都是相减混色。

相减混色就是以吸收三基色比例不同而形成不同的颜色的，所以又把青色、品红、黄色称为颜料三基色。颜料三基色的混色在绘画、印刷中得到广泛应用。在颜料三基色中，红绿蓝三种颜色被称为相减二次色或颜料二次色。在相减二次色中有：

青色＋黄色＋品红＝白色－红色－蓝色－绿色＝黑色

**4. HLS(色相、亮度、饱和度)原理**

HLS的含义是Hue(色相)、Luminance(亮度)、Saturation(饱和度)。

色相是颜色的一种属性，它实质上是色彩的基本颜色，即我们经常讲的红、橙、黄、绿、青、蓝、紫七种，每一种代表一种色相。色相的调整也就是改变它的颜色。

亮度就是各种颜色的图形原色(如RGB图像的原色为R、G、B三种或各自的色相)的明暗度，亮度调整也就是明暗度的调整。亮度范围从0到255，共分为256个等级。而我们通常讲的灰度图像，就是在纯白色和纯黑色之间划分了256个级别的亮度，也就是从白到灰，再转黑。同理，在RGB模式中则代表各原色的明暗度，即红绿蓝三原色的明暗度由浅到深。

饱和度是指图像颜色的彩度。对于每一种颜色都有一种人为规定的标准颜色，饱和度就是描述颜色与标准颜色之间的相近程度的物理量。调整饱和度就是调整图像的彩度。将一个图像的饱和度调为零时，则图像变成一个灰度图像，大家在电视机上可以试一试调整饱和度按钮。

另外还有一个概念，就是对比度。对比度是指不同颜色之间的差异。对比度越大，两种颜色之间相差越大，反之，就越接近。如一幅灰度图像提高它的对比度会更加黑白分明，调到极限时，变成黑白图像，反之，我们可以得到一幅灰色的画布。

## 9.6.3　三基色原理的应用

**1. 彩色电视系统**

景物光像(即包含各种颜色成分的自然光)分解为$R$、$G$、$B$ 3个分量(光像)，由摄像器件分别将有路光像转换为$E_R$、$E_G$、$E_B$ 3路光路电信号(其中，$E_R \propto R$、$E_G \propto G$、$E_B \propto B$，即电信号与光像的亮度成正比)，经过加工、传输与存储之后，在接收端将从接收信号中恢复的$E_R$、$E_G$、$E_B$ 3路光路电信号分别调制显像器件$R$、$G$、$B$3个电子枪的阳极，从而使得电子束的速度与$E_R$、$E_G$、$E_B$成正比，3束电子束分别轰击所对应3种荧光粉使其发光，因而，3种荧光粉的发光度与3路电信号成正比(即$E_R \propto R$、$E_G \propto G$、$E_B \propto B$)，于是荧光屏上每个像素的$R$、$G$、$B$荧光粉单元的发光亮度就与原图像对应像素的$R$、$G$、$B$成正比。由于荧光屏

上每个像素中的3个荧光粉单元的面积相对观看距离够小,因此,它们所发出的$R$、$G$、$B$ 3种色光同时刺激人眼时,就会在人眼的视网膜上产生的混色现象,从而使人眼能在荧光屏上观看到与摄像端景物光像相近的彩色图像。

根据色光三基色原理,人们采用如图9-6-3所示的系统来实现电视图像的传输。

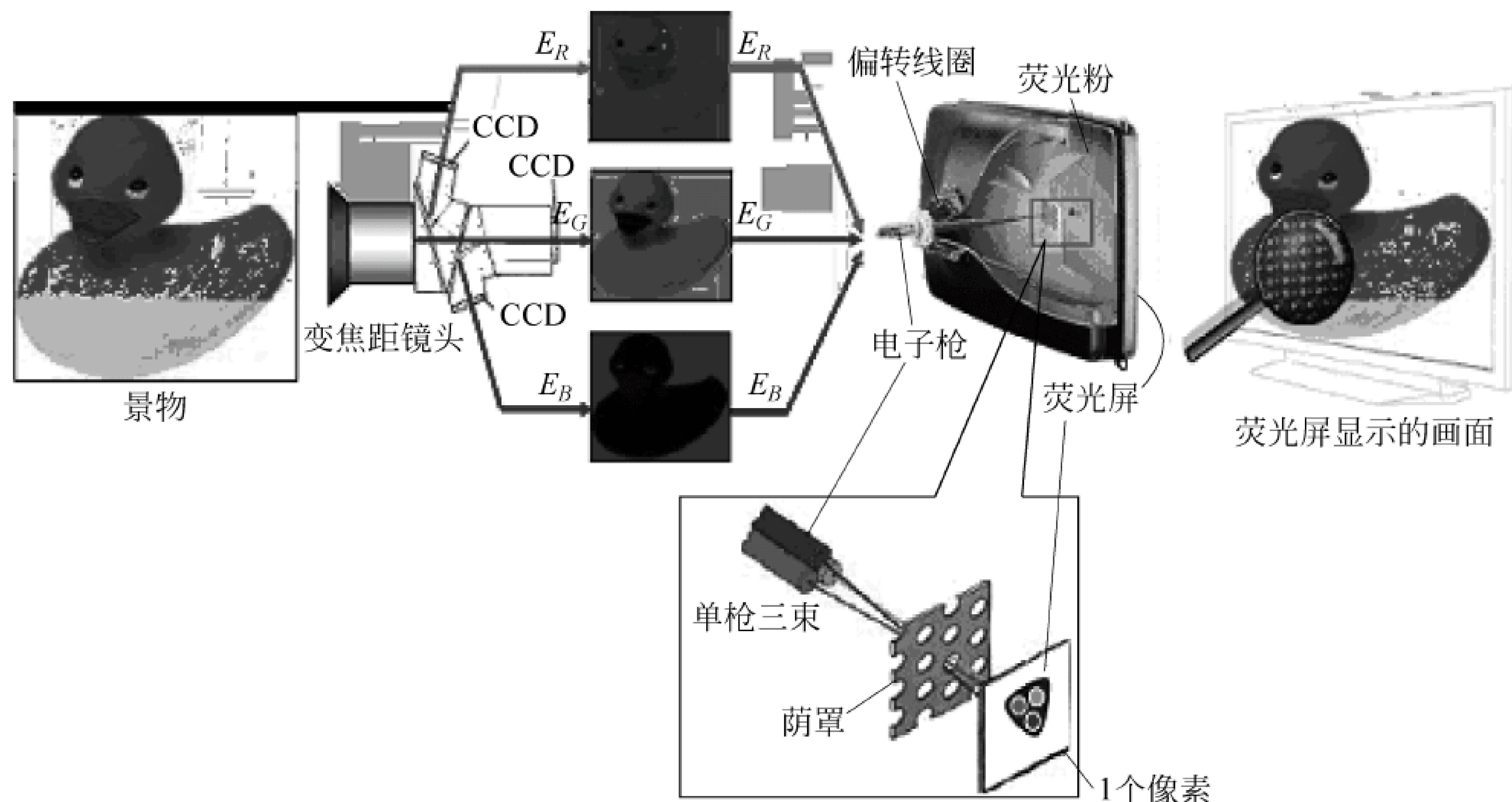



图 9-6-3 基于三基色原理的彩色电视系统

**2. 兼容制彩色电视系统**

所谓兼容制彩色电视系统是指出它所播出的彩色电视广播信号不仅能被彩色电视机接收到彩色图像,也能让黑白电视机正常接收和观看到黑白图像。这种兼容制彩色电视系统所播出的彩色电视信号并非分3路将RGB传送的,而是先从RGB中提取出亮度(Y)信号,然后,在色度C信号一起复合为1路彩色电视信号(称为复合视频信号)进行传送。黑白电视机接收到复合视频信号后,只需将其中的Y信号取出,就能正常观看,但只是黑白图像。而彩色电视机则需要同时将Y和C取出,并由它们最后还原为RGB,这样就能观看到彩色图像信号。

根据C信号的编码方式不同,可将兼容制彩色电视分为不同制式。目前,世界上的兼容制彩色电视主要有PAL制式、NTSC制式和SECAM制式3大类。

其中,采用NTSC制的国家主要有美国、加拿大等大部分西半球国家,以及中国的台湾、日本、韩国、菲律宾等采用这种制式;采用PAL制的国家主要有德国、英国等一些西欧国家,以及中国大陆和香港地区、朝鲜等国家采用这种制式;采用SECAM制的国家主要有法国、希腊、埃及、俄罗斯、东欧等。

## 9.7 衍射与全息照相

全息照相是一种新型的激光照相技术。“全息”是指物体发出的光波的全部信息:既包括振幅或强度的信息,也包括相位的信息。1948年伦敦大学的丹尼斯·伽柏(Dennis Gabor,1900—1979)首先提出了记录光波振幅和相位信息的照相原理,它和普通的摄影原理完全不同。后来人们把这种记录光波全部信息的照相技术称为全息照相。直到10多年后,发明了激光,全息摄影才得到实际应用。可以说,全息摄影是信息储存和激光技术结合

的产物。

### 9.7.1　全息照相与传统照相的区别

普通摄影(照相、电影、电视)是把摄影机的镜头对着被拍摄的物体,让物体上反射的光进入镜头,在感光底片上产生物体的像,感光底片上记录的是从物体上各点反射出来的光的强度,即记录的是来自物体的光,和物体有直接的关系。因为光是一种电磁波,决定波动特性的参数是振幅和相位:振幅表示光的强弱;相位表示光在传播中各质点所在的位置及振动方向。因此光的全部信息包括振幅和相位两个参数。普通摄影只是记录了光波的强度(正比于振幅的平方)信息,却没有记录光波的相位信息,它仅仅是一种平面影像。而全息照相不但记录了入射光波的强度,也记录了入射光波的相位。它是以干涉、衍射等波动光学理论为基础的无透镜拍摄,底片上所记录的是物体所发光波的全部信息(包括振幅和相位),因而可以再现物体逼真的立体形象。同时全息图中的每一个局部都包含了物体整体的光信息,即使底片有缺损,也不会影响完整物像的再现。

### 9.7.2　全息照相原理

人眼看见一个亮点,是因为有一个发光点所发出的球面波的波面被人眼接收到的缘故。如果发光的物体(可以看作是由无数点光源组成)被障碍物挡住,但它们所发出的球面波或特定的波面却被记录下来或被人眼看到,我们也应该感觉到该物体的存在,这就是全息照相的最初设想。事实上全息照相包含着两个内容:波前记录与波前再现。波前记录由物体反射(或透射)的光波(物光波)与另一参考光波相干涉,用感光底片将干涉条纹记录下来,形成全息图。波前再现,用一个与参考光波相似的光波照射全息图,光通过全息图产生衍射现象,衍射光波呈现出物体的再现像。

**1. 信息的记录**

简单来说,全息照片的记录原理就如图 9-7-1,将一束相干光(激光)垂直照射在两条平行狭缝 $S_1$ 和 $S_2$ 上,通过 $S_1$ 和 $S_2$ 的两束光,在屏幕 $D$ 上叠加成干涉条纹。如果把狭缝 $S_2$ 看作物体,$S_1$ 作为参考光束,则屏幕 $D$ 上的干涉条纹就是物体 $S_2$ 的全息图,用照相底片将它记录下来就得到一张狭缝 $S_2$ 的全息照片(它是一个明暗条纹的强度按正弦规律变化的光栅)。如果狭缝 $S_2$ 用其他实物替代,则可拍到实物三维空间的立体全息照片。

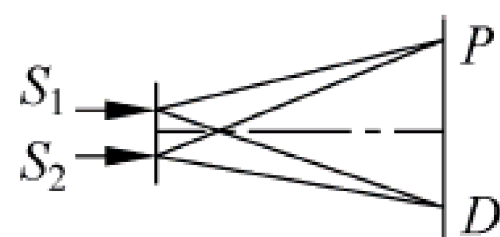


图 9-7-1　全息照片信息记录原理

详细解说如图 9-7-2,激光束经过分束镜后,分成两束相干光,一束足够强的相干光经反射镜反射,再经扩束后均匀地照射在被摄物体上,再从物体表面反射到感光板上,这束光称为物光。同时另一束相干光通过反射镜反射及扩束镜后直接投射到感光底板上,这束光称为参考光。物光与参考光在感光底板上叠加,发生干涉,感光底板经过曝光将这种图像记录下来,显影定影处理后就是一张全息照相的“照片”。从物体上各点反射出来的物光,其振幅和相位各不相同,所以感光片上各处的干涉条纹也不相同。振幅不同使条纹变黑程度不同;相位不同则使条纹的密度、形状各异。干涉图像的形状反映了物光与参考光束间的相位关系。而其明暗对比程度(称为反差)则反映了光波的强度(振幅的平方)关系,光束越强,明暗变化越显著,反差越大。由于物光的振幅和相位与物体表面各点的分布和漫射性质有关,从不同物点来的物光光程(相位)不同,所以干涉图像与被摄物有一一对应关系。因这种“照片”把物光波的全部信息都记录下来,故称为全息照相。

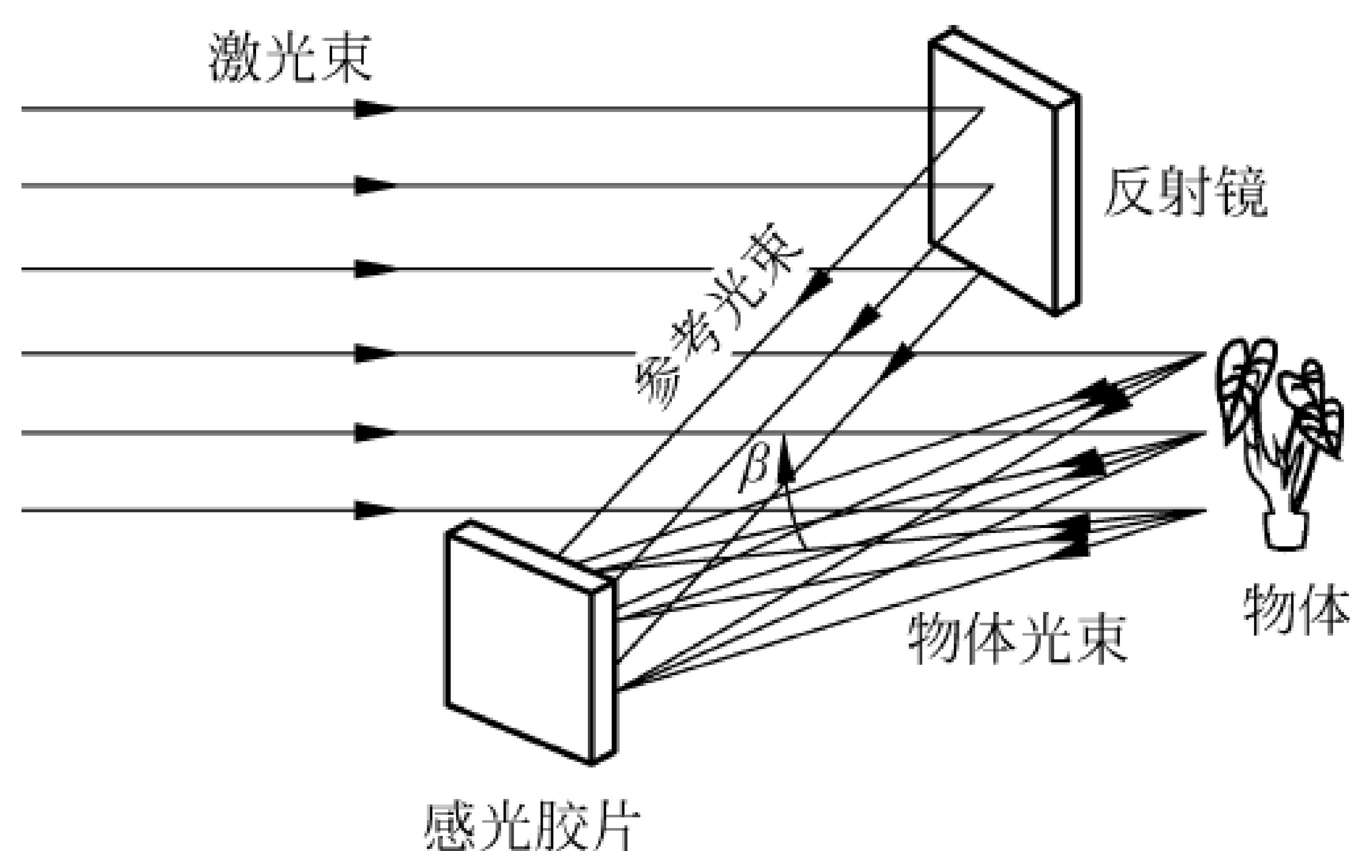



图 9-7-2 全息摄影原理

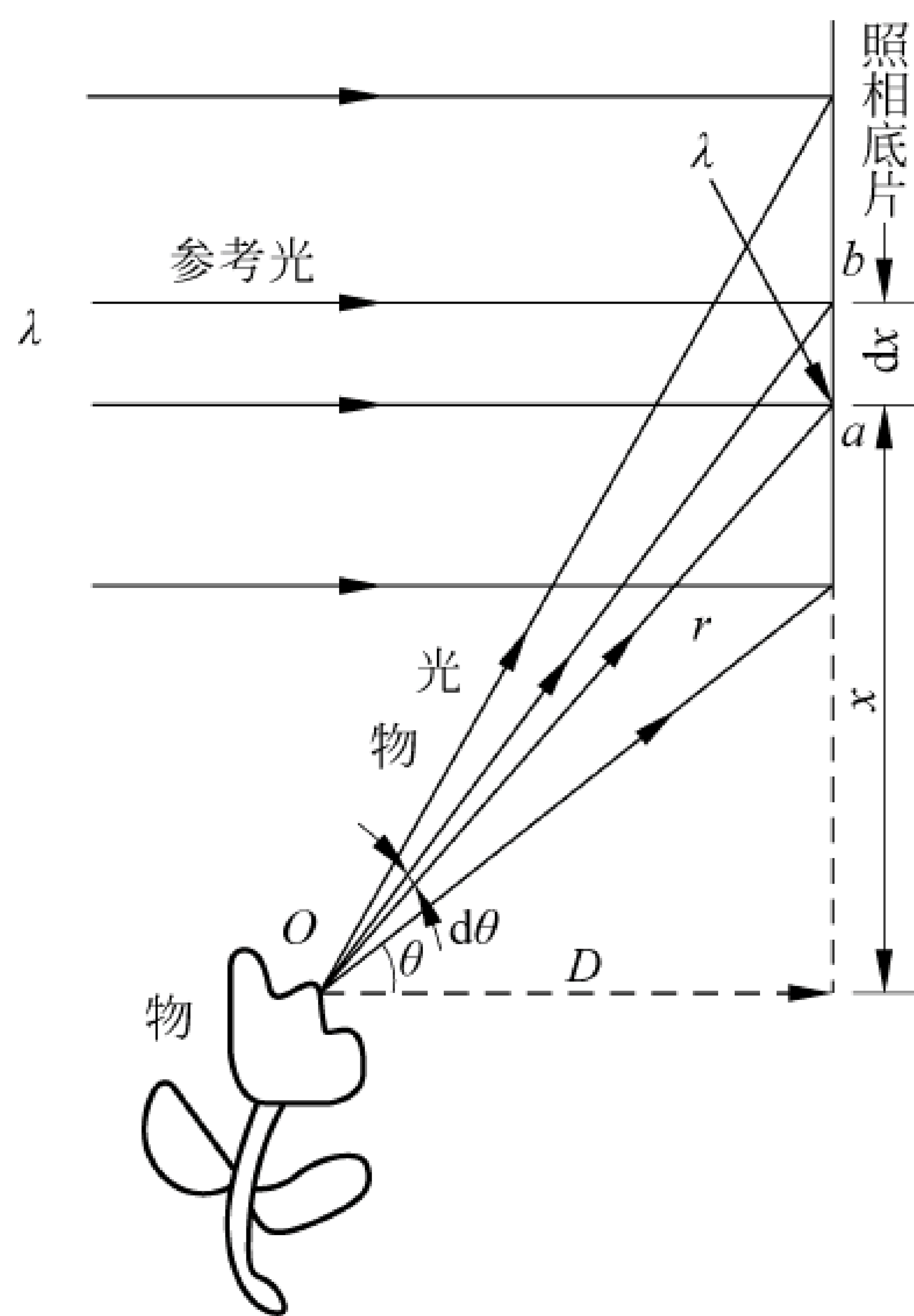



图 9-7-3 相位变化

此外，我们还可以研究全息照片的相位是如何变化的。如图 9-7-3 所示，设 $O$ 为物体上某一发光点，设参考光在 $a$ 处的振动方程为

$$y = A\cos(\omega t + \varphi_0) \tag{9-7-1}$$

物光在 $O$ 点处的振动方程为

$$y = A_1\cos(\omega t + \varphi_1) \tag{9-7-2}$$

物光在 $a$ 点处的振动方程为

$$y = A_1\cos(\omega t + \varphi_1 - 2\pi r/\lambda) \tag{9-7-3}$$

参考光与物光在 $a$ 处的相位差为

$$\delta\varphi = \varphi_0 - \varphi_1 + 2\pi r/\lambda \tag{9-7-4}$$

由干涉原理知 $\delta\varphi=(2k+1)\pi$ 处为暗条纹，解得

$$r = \lambda[(2k+1)\pi + \varphi_1 - \varphi_0]/2\pi \tag{9-7-5}$$

在 $\delta\varphi=2k\pi$ 处为明条纹，解得

$$r = \lambda(2k\pi + \varphi_1 - \varphi_0)/2\pi \tag{9-7-6}$$

设 $a$、$b$ 为相邻的两暗纹，由干涉知：$a$、$b$ 两处的物光与参考光必须都反相。因为 $a$、$b$ 两处的参考光相同，所以其物光的波程差为 $\lambda$，由几何关系知 $\lambda=\sin\theta d_x$ 和 $d_x=\lambda/\sin\theta$。

由此可知：当 $\theta$ 不同时，物光与参考光形成的干涉条纹的间距也不同，而 $\theta$ 的大小又可以反映出物光光波的相位，再根据条纹的方向即可确定出物体的前后、上下、左右的位置。

**2. 信息的再现**

详细来说，由于全息照相在感光底板上记录的不是物体的直观形象，而是无数组干涉条纹复杂的组合，所以观察全息照片记录的物像时，必须用与原来参考光完全相同的光束去照射，这束光称为再现光。用再现光观察时所用光路如图 9-7-4 所示，用一束被扩束的相干光从特定方向照射到全息照片上，当参考光束通过全息照片时，在它的后面出现一系列零级、一级、二级等衍射波。零级波可以看成是衰减后的入射光束。两个一级衍射波构成物体两个再现的像。一个为虚像，观察者在全息照片的后面可以看到；另一个为实像，观察者在全

息照片的前面可以看到。再现像与原来物体的形象完全一样。实像可以用感光胶片拍摄下来。

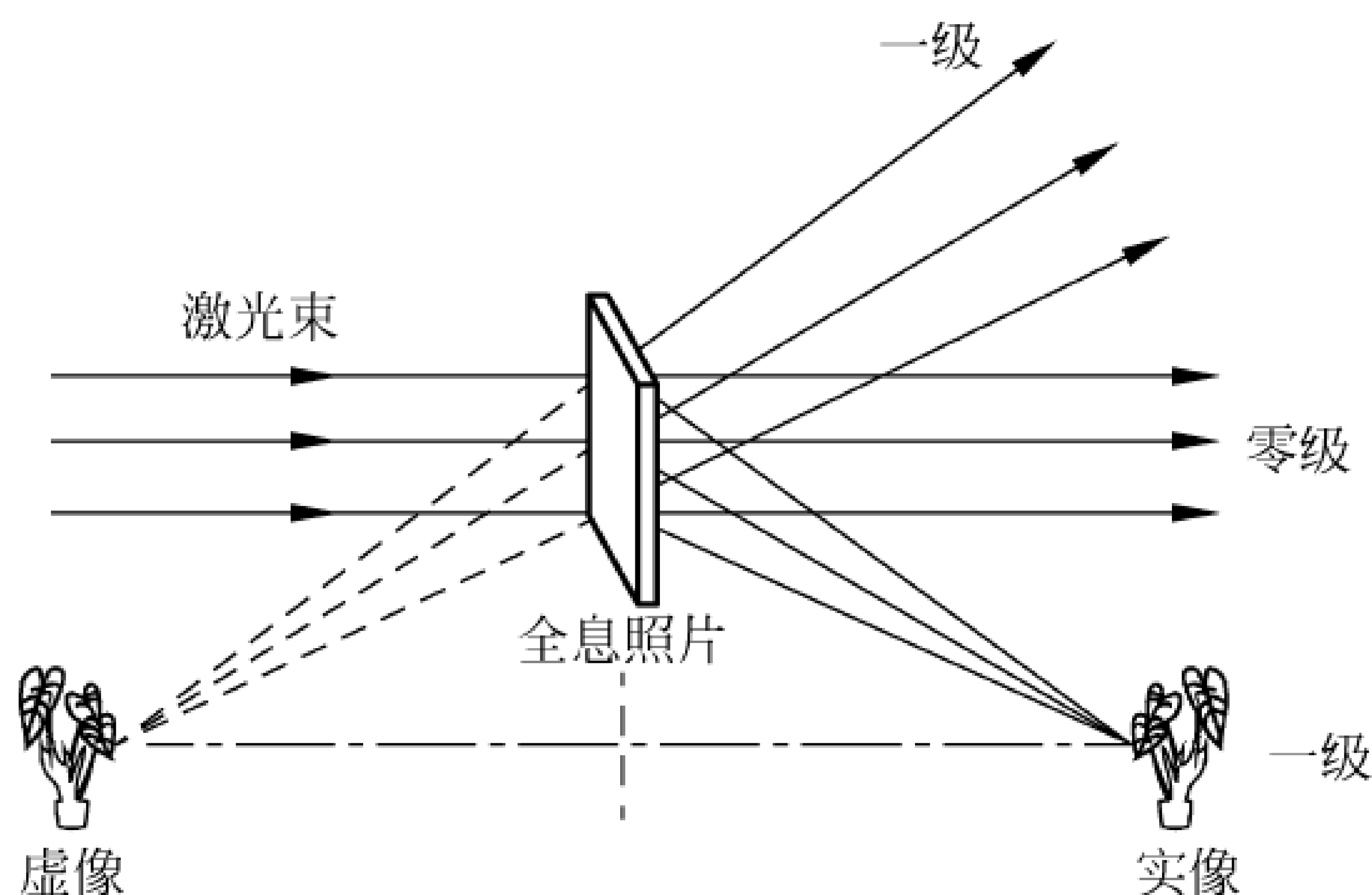



图 9-7-4　全息照片的再现原理

全息照片不同于普通照片,其底片不显示物体的形象,而是干涉条纹叠加后的图像。冲洗时只是改变了不同部分的透光性。所有发光点对应的衍射条纹会使人眼看到一个处于原来位置的完整的立体虚像。当人眼换一个位置观察时,会看到物体的侧面像,而且原来被其他物体遮住的地方这时也能显露出来。由于在拍摄时物体上任一发光点发出的物光在整个底片上各处都和参考光发生干涉,因而底片上各处都有该发光点的信息记录。所以,即使是取底片上的一小块残片来观察,也照样能看到整个物体的立体形象。

### 9.7.3　全息照相实验

全息照相术是在感光板上记录物体表面反射(或漫反射)光与参考光的干涉图样。干涉图样形成复杂光栅,在参考光独立照射干涉图样负片时,衍射形成物体的实像和虚像。全息照相目的就是要在实验的基础上了解全息照相的基本原理,掌握拍摄全息图的方法,通过实验了解全息照相的特点。实验中用到的实验仪器有全息台,He-Ne 激光器,分光镜,反射镜(两块),扩束镜(两个),被摄物体,曝光定时器,全息感光板,底板架,显影和定影液及相关器具,如图 9-7-5 所示。



图 9-7-5　全息照相的实验装置

**1. 静物全息照片的拍摄**

(1) 调整光路:使物光与参考光夹角约为 30°,分束器为透过率 96%的平板,以满足参、物光比为 1∶1～10∶1。具体调节分以下两步。

调节光学元件的螺钉,使光束基本同高,调节扩束镜的位置,使扩束后的光均匀照亮被摄物体。但光斑不能太大,以免浪费能量。在底片夹上放一张白纸,调节底片夹位置,使白纸上出现物体漫反射来的最强光。挡住物光,调节全反射镜,使参考光与物光中心反射到底片的光之间的夹角为 30°,并经扩束镜后,最强的光均匀地照亮底片夹上的白纸。

调整光程差 $D$ 等于零或近似为零,调节参考光的全反射镜,尽量使物光与参考光等光程,即用软质米尺(或细麻绳)从分束器量起,使物光光程等于参考光程。

(2) 曝光：调好光路后，打开曝光定时器，选择预定曝光时间，一般用1～2mW的激光管，定20s左右的曝光时间。让曝光定时器遮光，在全黑条件，取下底片夹白纸，装上干板(药面向被摄物体)，让环境稳定2～3min后，打开曝光定时器进行曝光(千万注意：此时切勿走动或高声谈话等)。

(3) 冲洗干板，将已曝光的干板取下(切忌手指触底片中间位置)，在全黑条件下，于D-19显影液中显影，待显至需要的时间，一般约5min(在暗绿色安全灯下观看，底片曝光部分呈现黑色斑纹即可)，取出干板，经水洗、定影、再水洗即可得到所需的漫反射三维全息照片。

定影时间为5～10min。D-19显影液显影，F-5定影液定影。显、定影时间也可由实验室提供。干板冲洗、晾干后即可观察干涉图样和再现像。

**2. 观察全息照片的再现物像**

将全息照片架在支架上，用参考光束作为照明光束照明全息照片，则可以观察到衍射的三维虚像，而且利用全息图上任何部分都能再现原物的整体像。

观察虚像后，将全息片绕铅直轴180°仍放在支架上，这时照明光束从全息图的背面照射，在全息图的前方用一白纸屏可看到物体的再现实像。

注意事项：

(1) 勿用手、手帕、纸屑等物擦拭光学元件。

(2) 曝光时切勿触及全息台，不要随意走动，防止实验室内有过大的气流流动。

(3) 不能用眼睛直视未扩束的激光束，手切勿触激光管高压端。

(4) 全息底板是玻璃片基，注意轻放以免弄碎。

### 9.7.4 全息照相技术的应用

**1. 全息照相的优越性**

由于全息照相具有独特的优点，应用潜力大，所以越来越受人们的重视。从其照相原理可以知道它具有以下主要特点：

(1) 由于全息照相记录了物体光波的全部信息，所以再现出来的物体形象就和原来的物体一模一样，是一个十分逼真的立体像。而且这种立体像还具有一些普通立体像所没有的优点：它和观察到的实物完全一样，具有相同的视觉效应。

(2) 全息照片的每一部分，无论有多大，总能再现出原来物体的整个图像。也就是说，可以把照片分成若干小块，每一小块都能完整地再现原来的物像。只是当全息照片的面积缩小后，像的分辨率降低了而已。全息照片的这一特点是由于照片的每一点都受到被摄物体各部分反射光的作用。所以全息照片即使有缺损，仍能再现被摄取的全部物像。

(3) 同一张底片上，经过多次曝光后，可以重叠许多像，而且每一个像又能不受其他像的干扰而单独地显示出来。如果对不同的景物采用不同入射角度的参考光束，由于所得到的干涉图样随物光和参考光之间的夹角大小而变化，因此相应的各种景物的再现像出现在不同的衍射方向上，因而在各个不同的地方组成了各个景物的独立的再现像。

**2. 全息照相技术的应用**

近年来全息照相技术已经有了很大的发展，现在能用普通光再现的全息图已经拍摄出来了。用这种方法可以制成彩色立体电视、彩色立体电影。但是，全息术的应用还很广，其很多用途还没开发出来。

（1）全息干涉测量

全息干涉测量是全息照相的主要应用之一，特别适用于各种材料的物损检验。全息干涉测量技术是把全息照相和干涉测量技术结合在一起形成的。其主要方法是先把被测物体经过全息照相，记录该物体的精确形象，如果物体在这之后发生了某种形变，可以把这个物体的形状与被记录的物体的原始形状进行比较，从而精确地指出物体形状的微小变化，进而判断物体的某些性质和形状。具体方法有一次曝光和二次曝光。

（2）全息显微术

由于全息照相的波前再现的这一特点，使得它与通常的成像技术相比更适宜用于显微技术。全息显微技术的放大，是借助于记录与重现过程中光波波长改变及曲率半径改变而实现的。全息显微镜最大的优点是扩大了显微镜的景深。另外，用全息显微技术无需把样品制成薄片，甚至对活的标本也可进行显微观测。加上重现的像具有立体性，能显示出样品的细节，从而可以加深对微观世界的认识。

（3）全息技术在海洋学中的应用

用激光全息技术进行水下观察，比起直接用光学以及声呐那样的常规搜索和监视技术优越得多。直接观察的距离近，而常规的声呐不能提供一个可辨认和分析的像。但是，激光全息却可以在较大的视野内获得水下物体的清晰的像。激光全息技术对于探测海中沉没物体、海底地形测绘、海洋资源考察、救生工作以及舰船导航和操纵潜艇在狭窄海峡内航行等都是十分有价值的。

（4）全息照相制作光学元件

用全息照相法制作光学元件能极大地改进光学元件的性能。例如，用全息法制作光栅完全摆脱了机械刻划的陈旧方式，而采用记录干涉条纹的照相方法就可以方便地得到质量较好的衍射光栅。全息光栅的特点是杂散光很小，且一般分辨率高。此外，全息图经常还可作为光学元件使用，如透镜、校正器、波带片等。

（5）全息信息储存

在拍摄全息照片时，改变参考光束的方向，可以将不同物体摄制在同一张底片上。再现时，只要偏转照明光束，就能将各物体互不干扰地显现出来。一张底片可以储存许多信息，如文字、图表或其他资料等，全息照片正在发展成为信息存储器，其存储量要比目前使用的其他存储器高一到两个数量级。

总的来说，全息技术再现出来的像是跟原来物体一模一样逼真的立体像，跟直接观察实物感觉完全一样。即使把全息照片分成若干小块，每一小块都可以完整地再现原来物体的像，所以全息照片即使有缺损，也不会使像失真。在同一张感光片上可以重叠记录许多像，这些像能够互不干扰地单独显示出来，而且全息照片很易复制。

## 9.8　菲涅耳公式与布儒斯特定律

光的干涉和光的衍射，证明了光是以波动形式来传播的。但是，波是振动的传播过程，如果以振动形式来分，波还可以分为纵波和横波两种。那么，光到底是横波还是纵波呢？这个问题，在光的干涉、衍射现象中不能得到结论，因为不论是横波还是纵波，只要是波动，都可以产生干涉和衍射现象。

### 9.8.1　光的偏振

光的偏振现象可以从实验上清楚地显示出光的横波特性，这一点和光的电磁波动理论

一致。可以说,光的偏振现象为光的电磁波本性提供了进一步的证据。

光的偏振态是指垂直于光线方向的二维平面上,光矢的运动状态。按光的偏振态,可以将光分为自然光和偏振光两大类。

普通光源(自发辐射为主导的光源)的大量发光原子中,各原子每一次辐射可以取不同的时间和不同的光振动方向,如果迎着光线方向看过去,如图 9-8-1(a)、(b)所示,光矢以相等的振幅均匀地分布在垂直于光传播的平面上。这些振动或者同时存在,或者迅速而无规则地替代,它的特点是光振动方向是随机的,但统计地说,相对光线方向是对称的。这种普通光源发出的光就叫作自然光。因为任何取向的光矢量都可以在两个互相垂直的方向上分解,因此自然光可以用两个互相没有稳定相位差,振动方向互相正交的,振幅 $A_x$、$A_y$ 相等的两个光振动代替,且有 $I_x=I_y=\frac{1}{2}I_0$。简言之,自然光可以看作是两正交方向振动的、没有固定相位差的、等振幅的线偏振光的混合。

完全偏振光指光矢的振动方位不变,或者有某种规则变化的光波,对完全偏振光细分,又可分为下面几种类型:线偏振光、圆偏振光、椭圆偏振光。

线偏振光又称平面偏振光,是指光矢量只沿一个固定方向振动的光。如果在垂直光线的平面上看,光矢轨迹为直线(图 9-8-2)。

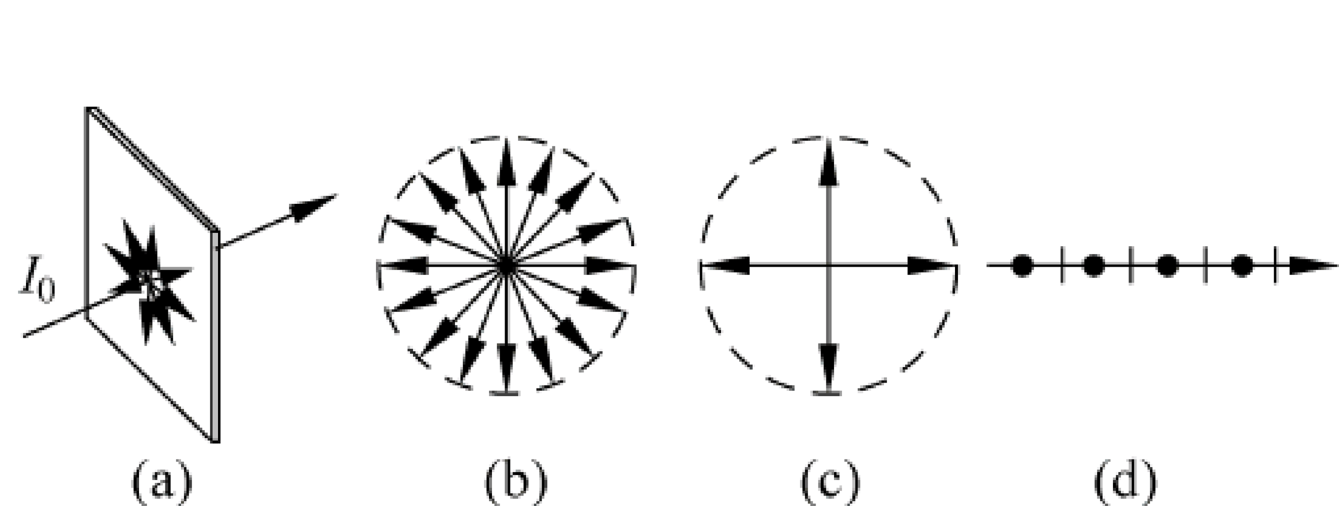



图 9-8-1 自然光

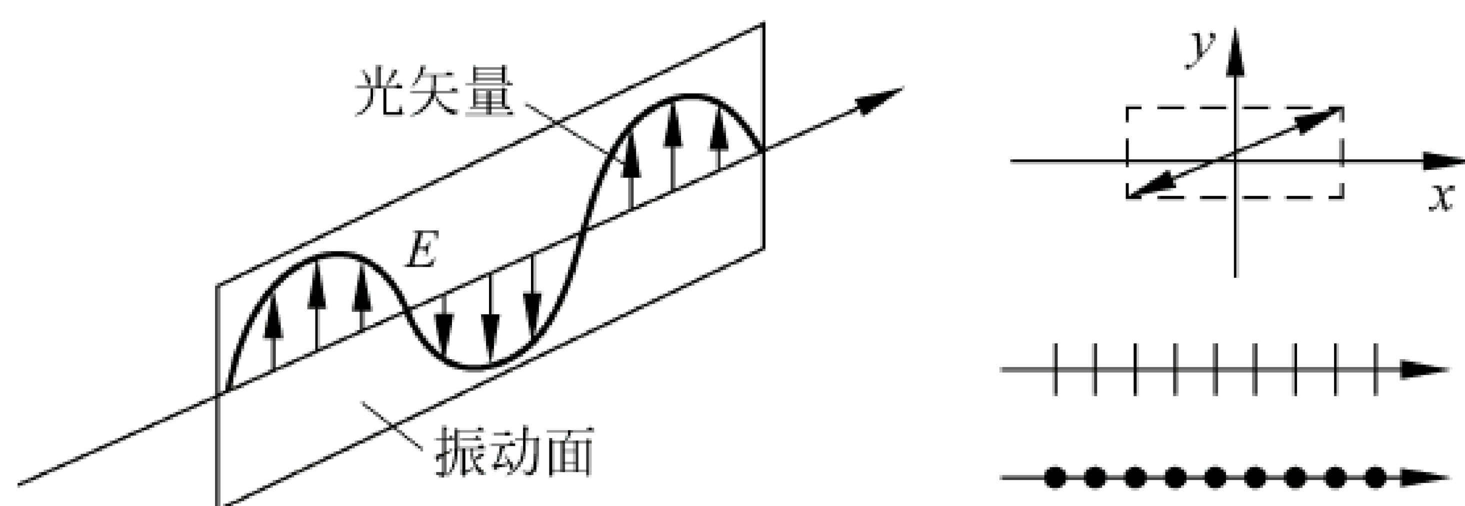



图 9-8-2 线偏振光

在垂直于光的传播方向的平面内,光矢量以一定的频率旋转。矢量端点轨迹为椭圆时称其为椭圆偏振光,如图 9-8-3(a);轨迹为圆时称其为圆偏振光,如图 9-8-3(b)。若在迎光的传播方向观察圆偏振光的光矢量随时间变化是右旋的,则这种圆偏振光叫作右旋圆偏振光(图 9-8-4);反之,叫作左旋圆偏振光。

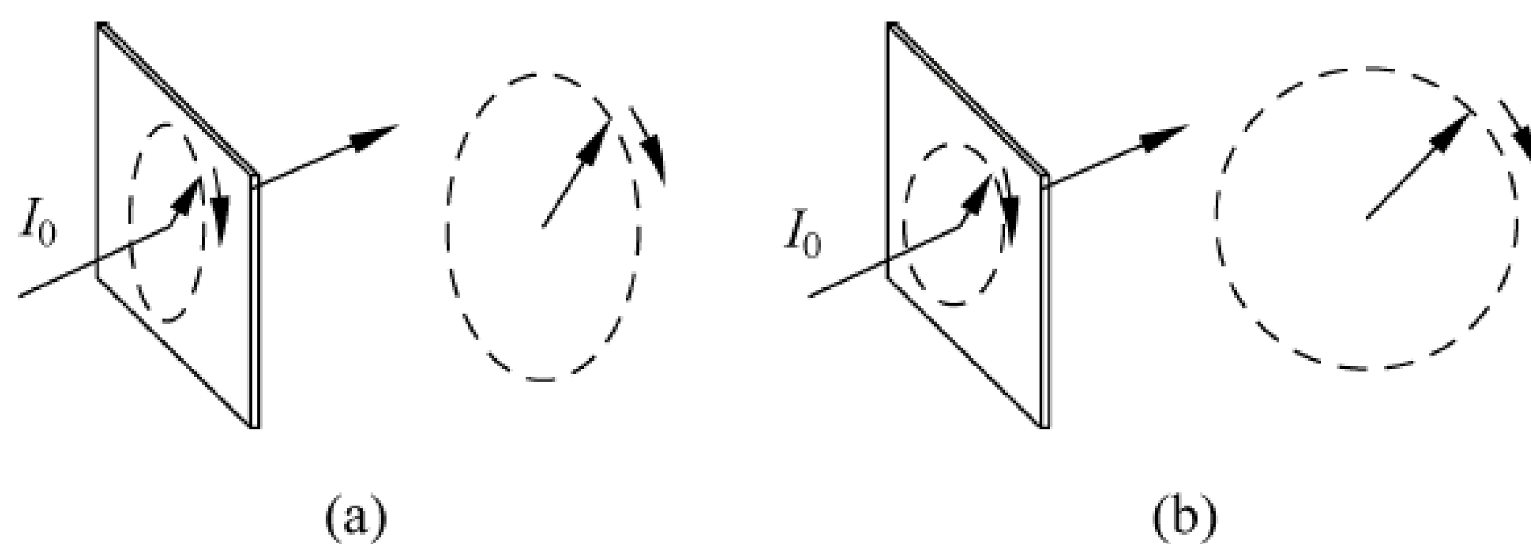



图 9-8-3 椭圆偏振光和圆偏振光

事实上,线偏振光、圆偏振光都可以看成是椭圆偏振光的特殊情况。

部分偏振光(图 9-8-5)是指在垂直于光传播方向的平面内,各方向都有光振动,但振幅不等的光。它是全偏振光和自然光的混合光。自然光和线偏振光的混合光叫作部分线偏振光;自然光和椭圆偏振光的混合光叫作部分椭圆偏振光;自然光和圆偏振光的混合光叫作部分圆偏振光。

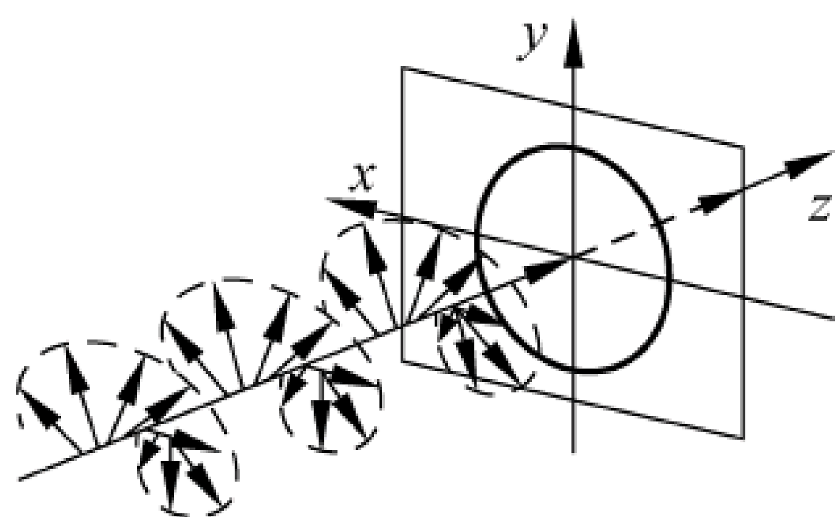



图 9-8-4　右旋圆偏振光

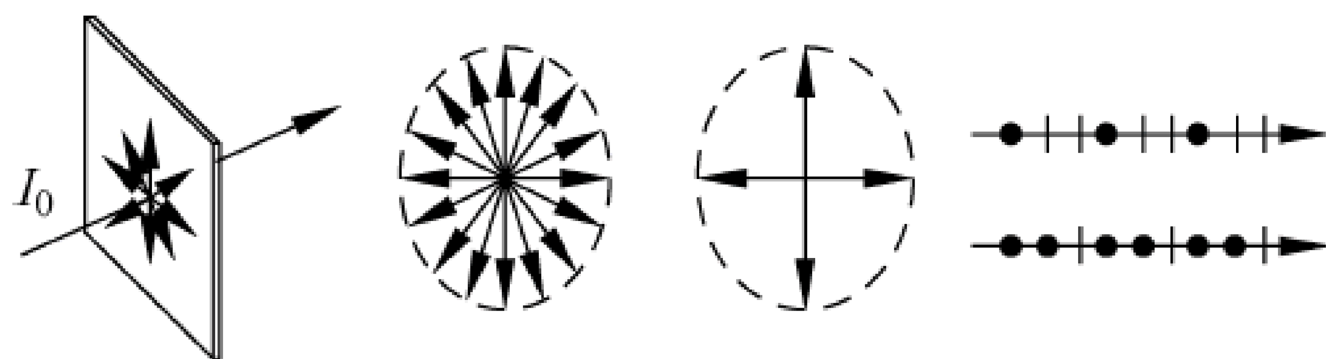



图 9-8-5　部分偏振光

## 9.8.2　菲涅耳公式

当光通过不同的介质界面时，入射光分为反射光和折射光两部分，折射定律和反射定律决定了它们的方向，而这两部分光的强度和振动的取向，则需要用电磁理论来讨论。

因为光在介质界面上反射和折射时，其平行于入射面的分量（又称平行分量，用 $p$ 表示，或简称 $p$ 分量）与垂直于入射面的分量（又称垂直分量，用 $s$ 表示，或简称 $s$ 分量）的行为不同。

自然光和偏振光都可以分解为两个相互垂直的线偏振光，所以讨论光在透明介质界面上所发生的现象时可以借助于讨论两个特定的线偏振光来进行（相当于 $p$ 分量和 $s$ 分量）。将入射光、反射光和折射光分解为 $p$ 和 $s$ 分量（图 9-8-6），根据电磁场的边界条件和振幅关系就推出菲涅耳公式。由于公式的数学推导较烦琐，这里直接给出结论。

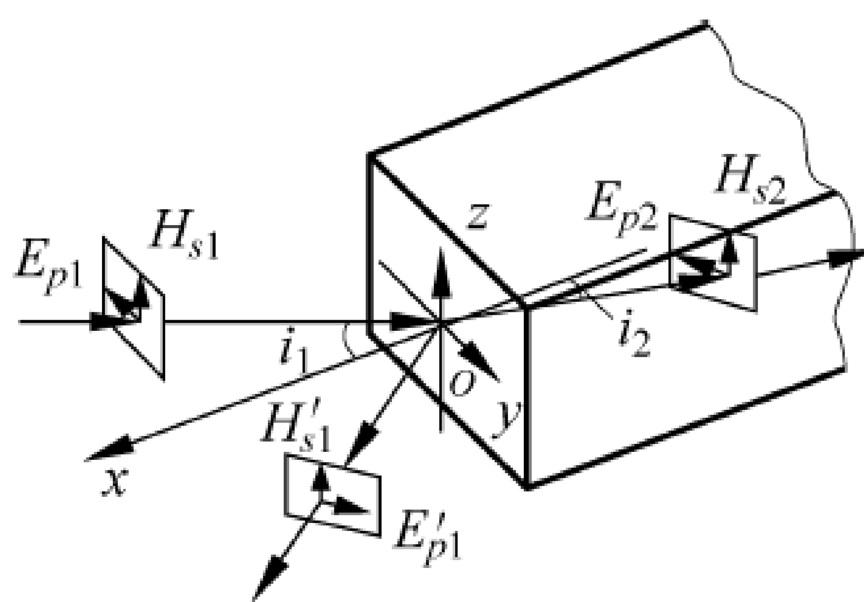



图 9-8-6　$p$ 分量和 $s$ 分量

$$r_s=\frac{E'_{s1}}{E_{s1}}=\frac{\sin(i_1-i_2)}{\sin(i_1+i_2)} \tag{9-8-1}$$

$$r_p=\frac{E'_{p1}}{E_{p1}}=\frac{\tan(i_1-i_2)}{\tan(i_1+i_2)} \tag{9-8-2}$$

$$t_s=\frac{E_{s2}}{E_{s1}}=\frac{2\sin i_2\cos i_1}{\sin(i_1+i_2)} \tag{9-8-3}$$

$$t_p=\frac{E_{p2}}{E_{p1}}=\frac{2\sin i_2\cos i_1}{\sin(i_1+i_2)\cos(i_1-i_2)} \tag{9-8-4}$$

前两式表示反射波的两个分量振幅和入射波的两个分量振幅之比；后两式表示折射波和入射波两个对应分量振幅之比。

振动方向的变化则由正负号来决定。各分量振幅量值之比是相对于入射波来计算的振动方向分别按照各波的上述规定，而不是直接相对于入射波作比较。

## 9.8.3　半波损失的条件

### 1. 正入射（$i\approx0°$）的半波损失问题

光从光疏介质射到光密介质界面反射时，因折射率 $n_2>n_1$，所以 $i>r$，$i+r<90°$。

由菲涅耳公式（9-8-1）和（9-8-2）得 $\frac{E'_p}{E_p}>0$，即 $E'_p$ 与 $E_p$ 符号相同；$\frac{E'_s}{E_s}<0$，即 $E'_s$ 与 $E_s$ 符号相反。

$E'_p$与$E_p$ 符号相同，也就是说，以光行进方向作参考，矢量 $E'_p$与$E_p$ 方向相同。如图 9-8-7 所示，无论对入射光还是反射光，$E'_p$与$E_p$ 的方向均指向前进方向的左边。但在参考系看来，$E'_p$与$E_p$ 的方向恰好相反。同理，可得 $E'_s$与$E_s$ 方向，如图 9-8-7 所示。

可见，无论是垂直入射面的电矢量还是平行入射面的电矢量，当 $n_2>n_1$，且 $i\approx 0°$时，反射电矢量与入射电矢量的相位差将有 π 的突变，即所谓半波损失。

但当光从光密介质射到光疏介质界面反射时，同上方法可知 $E'_p$，$E_p$，$E'_s$与$E_s$ 的方向，显然，反射电矢量与入射电矢量的位相差为零，没有半波损失。

**2. 掠入射时($i\approx 90°$)的半波损失问题**

若 $n_2>n_1$，所以 $i>r$，$i+r<90°$，由菲涅耳公式(9-8-1)和式(9-8-2)得$\dfrac{E'_p}{E_p}<0$ 和 $\dfrac{E'_s}{E_s}<0$。

以光行进方向作参考，$E'_p$，$E_p$，$E'_s$与$E_s$ 的方向如图 9-8-8 所示。由图可知，在实验室坐标系看来，反射电矢量与入射电矢量(无论平行入射面的振动还是垂直于入射面的振动)的位相差为 π，存在半波损失。

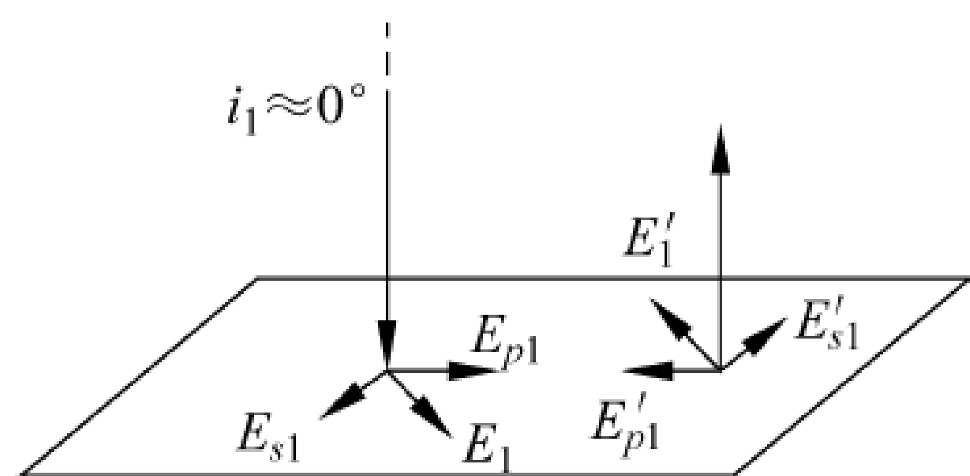



图 9-8-7 正入射的半波损失

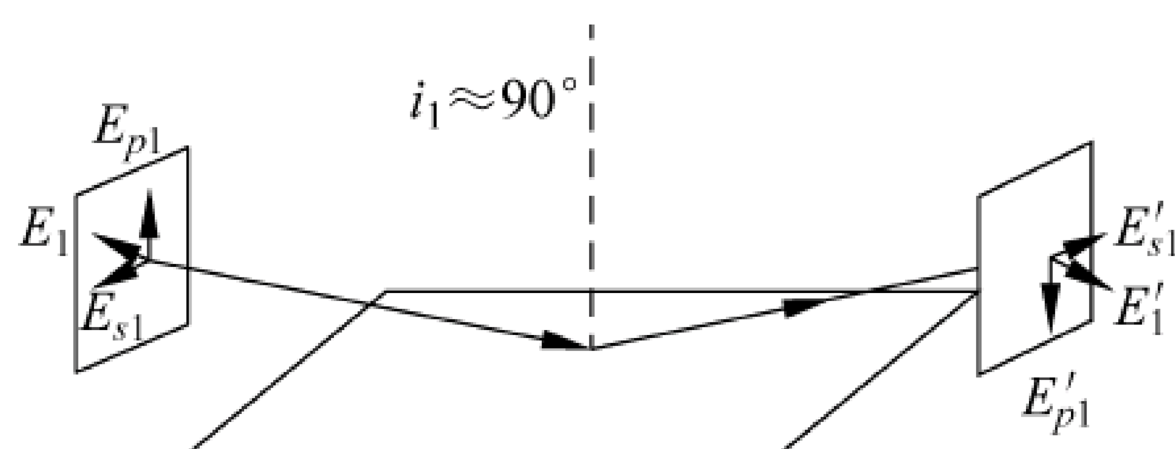



图 9-8-8 掠入射的半波损失

若 $n_1>n_2$，用同样方法可证明反射电矢量与入射电矢量的相位差为零(即没有半波损失)。

**3. 斜入射($0<i<90°$)的半波损失问题**

这时需要分四种情况进行讨论：

(1) $n_2>n_1$，且 $i<i_0$，即 $i+r<90°$

(2) $n_2>n_1$，且 $i>i_0$，即 $i+r>90°$

(3) $n_1>n_2$，且 $i<i_0$，即 $i+r<90°$

(4) $n_1>n_2$，且 $i>i_0$，即 $i+r>90°$

由菲涅耳公式可得四种情况下入射电矢量与反射电矢量之间的关系如图 9-8-9 所示。

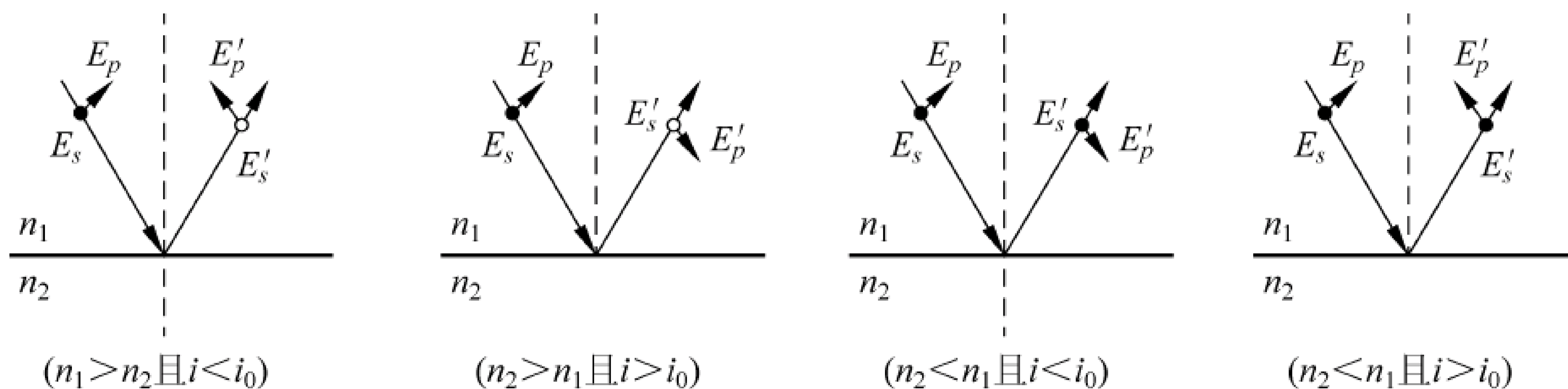



图 9-8-9 斜入射的半波损失

由图 9-8-9 可看出，对于平行于入射面的电矢量 $E'_p$与$E_p$，要比较它们的方向是相同或相反是毫无意义的。因而也不存在反射电矢量与入射电矢量的相位是否差 π 的问题。这种

情况下要谈半波损失问题也就没有意义了。

**4. 对几个光学实验的分析**

(1) 牛顿环实验

一般做牛顿环实验时，光是垂直入射的，从反射光观看牛顿环，其中心是暗斑，表明有半波损失。根据前面的结论可知，当光垂直入射时，光由光疏媒质到光密媒质的反射中，反射光有半波损失。实际的实验结果和以上的分析是一致的。

(2) 薄膜干涉实验

根据前面的分析，当入射光斜入射时，光在上下界面的反射都不存在半波损失。下面分析光经上下界面反射后，两束光的关系。根据菲涅耳公式可以推导出：

$$\frac{E'_{3p}}{E_{2p}}=-\frac{4\cos i_1\sin i_1\cos i_2\sin i_2}{\sin^2(i_1+i_2)\cos^2(i_1-i_2)} \tag{9-8-5}$$

$$\frac{E'_{3s}}{E_{2s}}=-\frac{4\cos i_1\sin i_1\cos i_2\sin i_2}{\sin^2(i_1+i_2)} \tag{9-8-6}$$

因为 $\left|\frac{E'_{3p}}{E_{2p}}\right|\neq\left|\frac{E'_{3s}}{E_{2s}}\right|$，两束反射光的 $p$ 分量和 $s$ 分量不成比例，即使 $E_{3p}$ 与 $E_{2p}$ 反向，同时 $E_{3s}$ 与 $E_{2s}$ 反向，反射光 $E_3$ 与反射光 $E_2$ 的振动面也不可能在同一个平面内，所以也就不存在传统意义上的“半波损失”。可实际上薄膜干涉公式中有 $\pm\lambda/2$ 项，表明有半波损失，且与实际观测结果相符。进一步分析可知，形成薄膜干涉的原因是，反射光 3 中与反射光 2 相平行的分量 $E''_3$ 与反射光 $E_2$ 产生干涉，$E''_3$ 的相位与 $E_2$ 的相反，可理解为多走了半个波长，即发生了“半波损失”。因此，薄膜干涉公式中的 $\pm\lambda/2$。应理解为是对 $E''_3$，而不是对整个 $E_3$ 而言的。因为 $E_2$ 与 $E_3$ 之间有夹角，所以会降低干涉条纹的反衬度。

(3) 洛埃镜实验

在一平面镜上方接近于平面镜处放一点光源，直接从点光源发出的光和由平面镜反射的光在空间相遇形成干涉，这就是洛埃镜实验。实验结果表明，当垂直于平面镜的接收屏与平面镜接触时，在接触处出现暗条纹，说明反射光有半波损失。

下面分析一下其产生半波损失的原因。如图 9-8-10 所示，从实验装置上看，要观测到明显的干涉条纹，要求光源 $S$ 很接近平面镜所在平面，这时点光源在平面镜上的入射角接近于 90°，即掠射。根据前面的分析可知，当光掠射时，光由光疏媒质到光密媒质的反射中，$N$ 处是暗条纹，说明反射光有半波损失。所以洛埃镜实验之所以可观察到半波损失，是因为入射光满足掠射的条件。

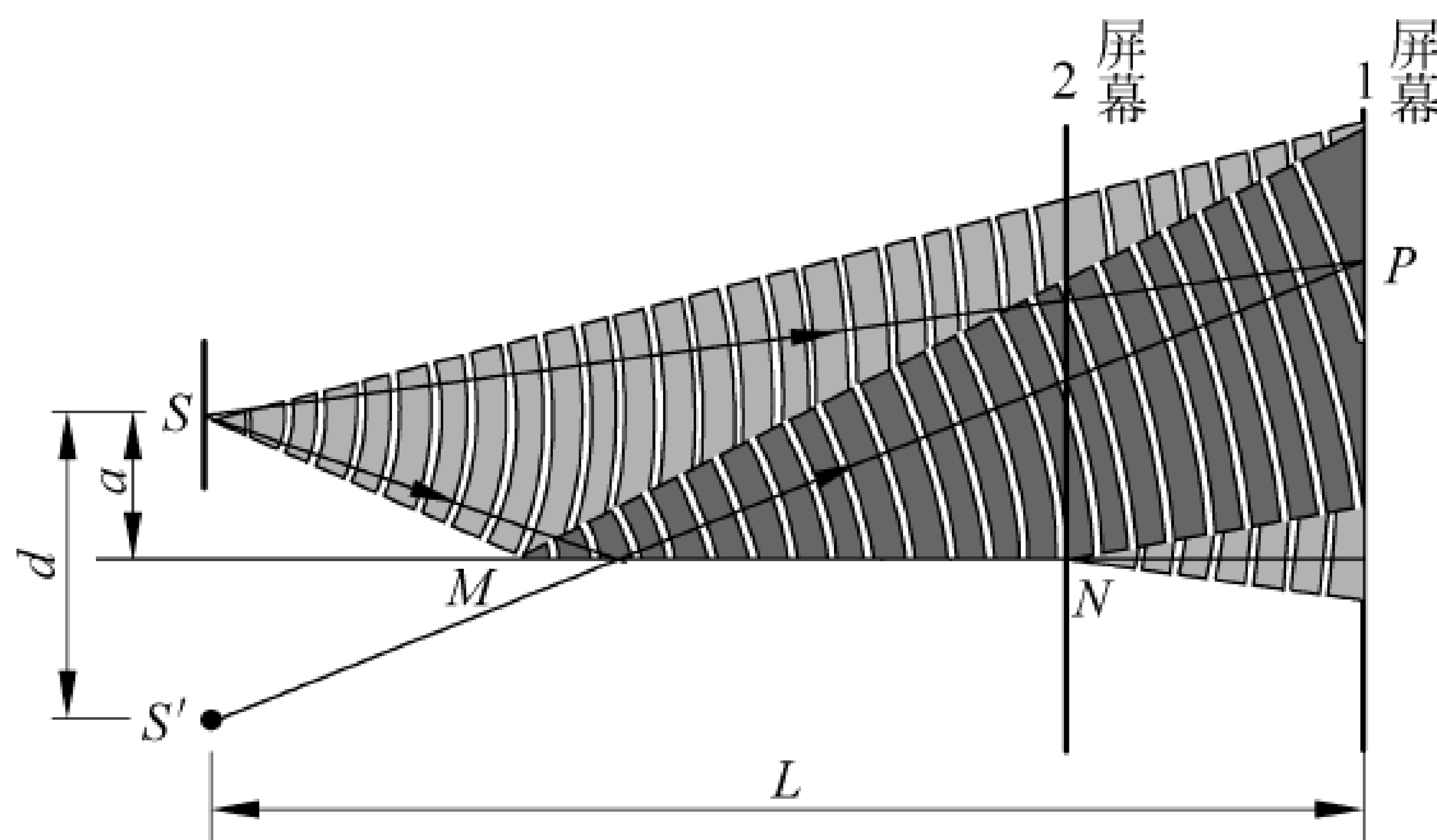



图 9-8-10　洛埃镜实验的半波损失

**5. 结论**

实验和理论表明，光在两种媒质分界面传播时，半波损失的产生与介质的折射率和入射角都有关系，可以得到反射光相对于入射光产生半波损失的条件是：

(1) 入射光由光疏媒质向光密媒质传播；

(2) 入射角 $i_1 \approx 0$ 或 $i_1 \approx \frac{\pi}{2}$。

## 9.8.4 布儒斯特定律

**1. 布儒斯特定律**

马吕斯(Etienne Louis Malus，1775—1812)曾发现，当一束光入射到玻璃板上时，反射光是部分偏振光，并且当入射角大约为57°时，反射光具有最大的偏振度，后来布儒斯特(David Brewster，1781—1868)发现折射光也是部分偏振光。用电磁理论也可以证明，光在两种介质的界面上发生反射和折射时，反射光和折射光确实都是部分偏振光。现在用检偏器也很容易对此进行检验。

总之，理论和实验都表明：自然光照射在介质表面发生反射和折射时，反射光与折射光一般来说都是部分偏振光。反射光中垂直入射面的光振动占优势，在折射光中则平行于入射面的振动占优势，如图9-8-11所示。

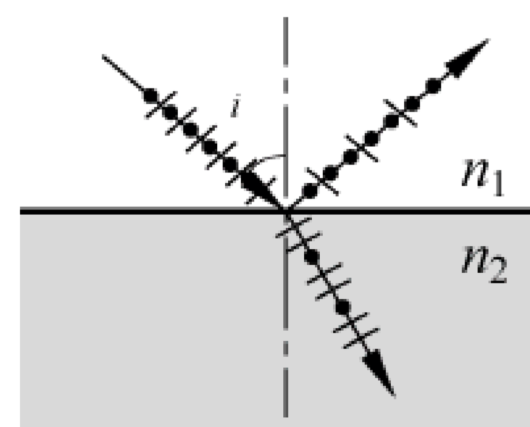


图 9-8-11 反射和折射时的偏振现象

改变入射角时，反射光和折射光的偏振度也发生变化。当入射光 $\theta_i$ 等于某一角度 $\theta_b$ 时，反射光为线偏振光。布儒斯特在1811年根据实验总结出一条定律，称为布儒斯特定律。其内容如下：对于一定的介质分界面，当入射角 $\theta_i=\theta_b$，而满足式(9-8-7)时

$$\tan i_b=\frac{n_2}{n_1} \tag{9-8-7}$$

反射光是线偏振光，它的振动方向和入射面垂直(图9-8-11)。式(9-8-7)称为布儒斯特角或全偏振角。

普通玻璃的折射率为1.54左右，空气折射率接近1，根据式(9-8-7)可以求出，当自然光从空气射向玻璃时布儒斯特角约57°。

布儒斯特定律有一条有用的推论：当入射角为起偏角时，反射光线和折射光线互相垂直。

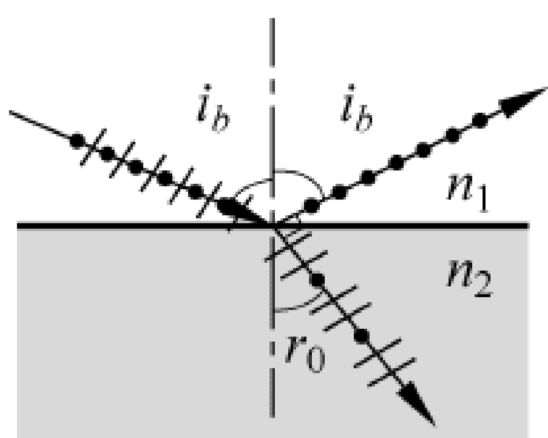


图 9-8-12 布儒斯特定律

证明如下，如图9-8-12所示，由折射定律可知：

$$\frac{\sin i_b}{\sin r_0}=\frac{n_2}{n_1} \tag{9-8-8}$$

由布儒斯特定律可知：

$$\tan i_b=\frac{\sin i_b}{\cos i_b}=\frac{n_2}{n_1} \tag{9-8-9}$$

于是有

$$\sin r_0=\cos i_b \tag{9-8-10}$$

最后得到

$$i_b + r_0 = \frac{\pi}{2} \tag{9-8-11}$$

命题得证。

关于布儒斯特定律，大家请注意：

（1）当入射角为布儒斯特角时，反射光为振动方向垂直入射面的线偏振光，而折射光仍为振动方向平行于入射面的成分占优势的部分偏振光。

（2）要注意布儒斯特角与全反射角的区别。

① 两者条件不同。全反射时对 $n_1$、$n_2$ 有要求，而布儒斯特角无此要求；

② 入射角大于全反射角时都会发生全反射，但只有入射角为布儒斯特角时反射光才是完全线偏振光。

**2. 布儒斯特定律的应用**

（1）玻璃片堆起偏

利用布儒斯特定律条件下的反射光，可以得到线偏振光，但往往光强较小；折射光的光强足够高，但偏振度太低。为了解决这个矛盾，可以利用玻璃片堆来提高透射光的偏振度和增加反射光的光强。

如图 9-8-13 所示，让自然光通过一个由多片玻璃叠合而成的片堆，并使入射角等于布儒斯特角。这样经过多次的反射和折射，可以使折射光有很高的偏振度，并且光强也可以达到入射自然光的 50%左右。

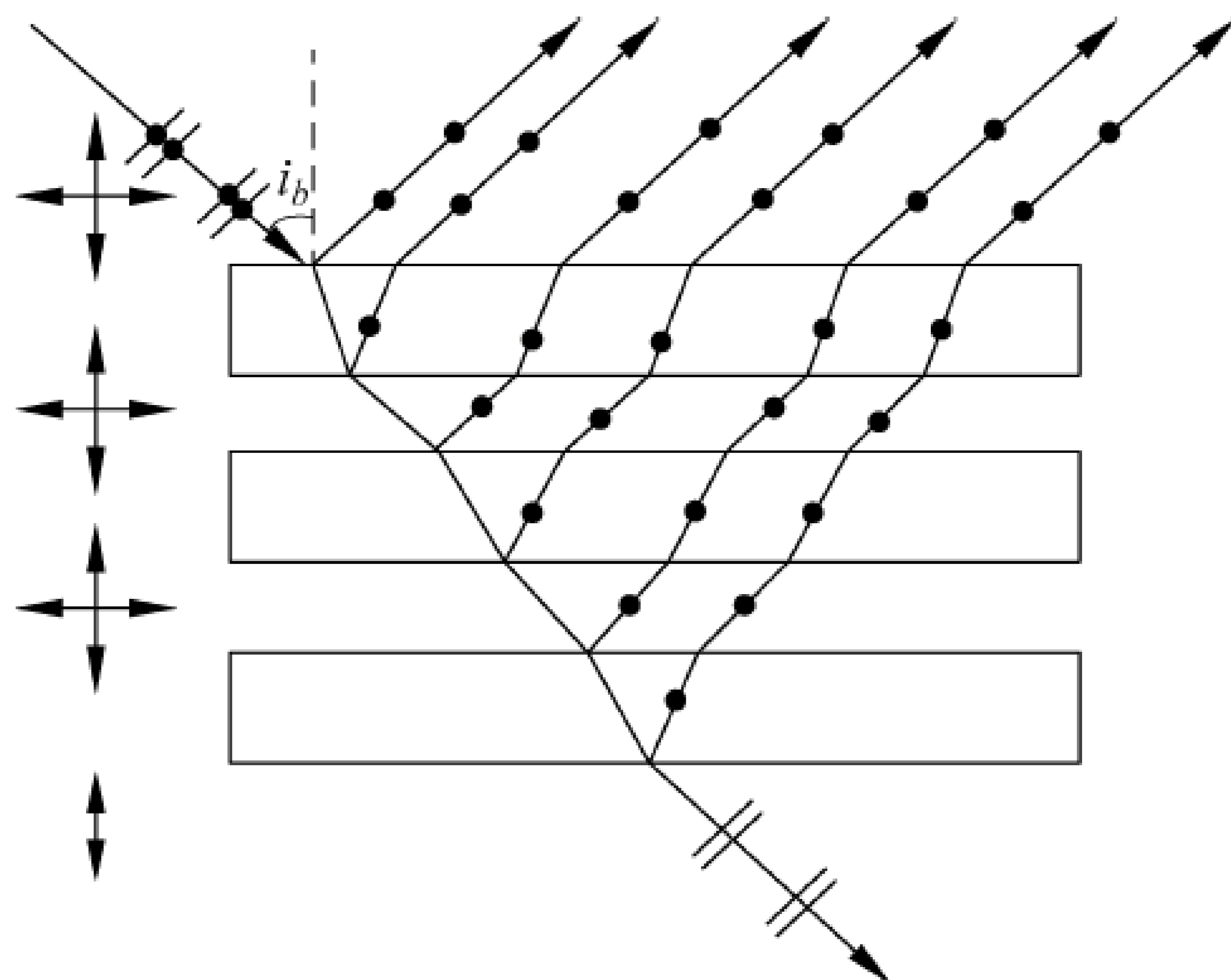


图 9-8-13　玻璃片堆起偏

（2）布儒斯特窗

利用布儒斯特定律还可以产生线偏振的激光。例如，在外腔式气体激光器中（图 9-8-14），激光管两端的透明窗（称为布儒斯特窗）$B_1$、$B_2$ 就是安置成使入射光的入射角成为布儒斯特角。在这种情况下，光矢量垂直于入射面的光（又称 $s$ 波），在一个窗上的一次反射损失约占 $s$ 波的 15%，虽然 $s$ 波在激光管内会得到能量补充，但由于损失大于增益，所以激光器谐振腔（反射镜 $M_1$、$M_2$ 之间的腔体）不能对 $s$ 波起振。而对于光矢量平行于入射面的光波（又称 $p$ 波），它在布儒斯特窗上没有反射损失，因而衰减很小，可以在腔内形成稳定的振荡，并从反射镜射出。这样，外腔式气体激光器输出的激光就是线偏振光。

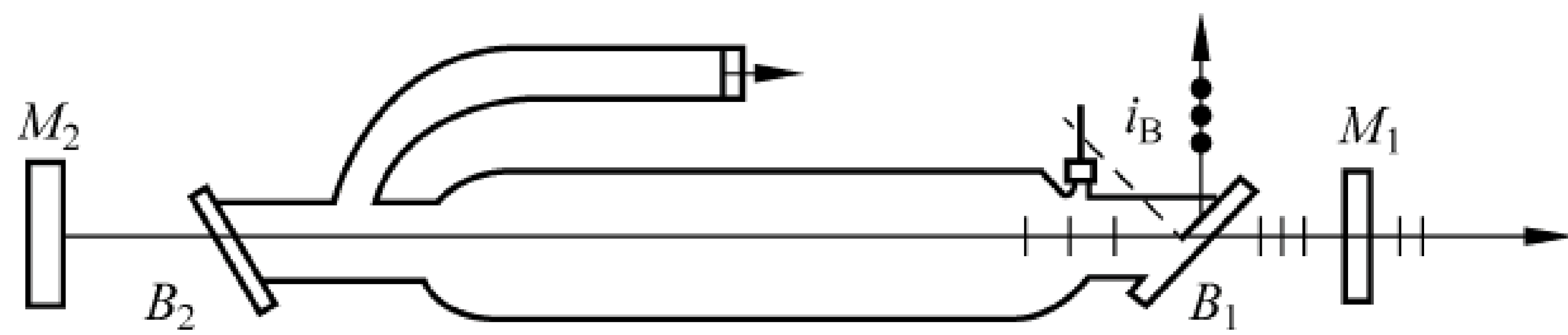


图 9-8-14　布儒斯特窗

# 学习资源

## 一、中国近现代著名科学家（物理学）简介

钱三强（1913 年 10 月 16 日—1992 年 6 月 28 日），浙江湖州人，原名钱秉穹，核物理学

家，中国科学院院士。



钱三强毕业于清华大学，巴黎大学。钱三强是中国原子能事业的主要奠基人和组织领导者之一，中科院学部委员、中科院特邀顾问、中国科协名誉主席、七届全国政协常委。他与妻子一同被西方称为“中国的居里夫妇”，他是中国发展核武器的组织协调者和总设计师，人称他领导的研究所“满门忠烈”。1999年9月18日，钱三强被追授“两弹一星”功勋奖章。钱三强是一位学识渊博、治学严谨、卓有成就的科学家，也是一名立场坚定的无产阶级战士，他艰辛开拓，毕生致力于祖国的科学技术事业，特别是为中国原子能科学事业的创立和“两弹”研制，为中国科学院及中国科学院学部的建立与发展，为培养和吸引人才，为国际学术交流，做出了卓越贡献。钱三强和钱骥、钱学森被并称为中国科技界的“三钱”。

钱三强夫人何泽慧(1914年3月5日—2011年6月20日)，女，核物理学家。1914年生于江苏苏州，籍贯山西灵石。1932年，何泽慧从外祖母创办的苏州振华女校高中毕业，随同学前往上海考大学，考入清华大学物理系。1936年毕业于清华大学。1940年获德国柏林高等工业大学工程博士学位。1946年春天，何泽慧从德国到法国巴黎，和大学时期的同学钱三强结婚，开始共同的科学生涯。他们一起在约里奥·居里夫妇领导的法兰西学院原子核化学实验室和居里实验室工作，合作发现了铀核裂变的新方式——三分裂和四分裂现象(她首先捕捉到世界上第一例四分裂径迹)，在国际科学界引起很大反响，所以其被西方媒体称为“中国的居里夫人”。何泽慧于1980年当选为中国科学院学部委员。中国科学院高能物理研究所研究员。在德国海德堡皇家学院核物理研究所工作期间，首先发现并研究了正负电子几乎全部交换能量的弹性碰撞现象。在法国巴黎法兰西学院核化学实验室工作期间，与合作者一起首先发现并研究了铀的三分裂和四分裂现象。20世纪50年代，与合作者一起自力更生研制成功对粒子灵敏的原子核乳胶探测器。在领导建设中子物理实验室、高山宇宙线观察站，开展高空气球、高能天体物理等多领域研究方面做出了重要贡献。何泽慧是中国人民政治协商会议第五、六、七届全国委员，空间科学学会原常务理事，中科院高能所原副所长，杰出的科学家。

钱伟长(1912年10月9日—2010年7月3日)，江苏无锡人，中国近代力学之父，世界著名的科学家、教育家，杰出的社会活动家，中国民主同盟的卓越领导人，中国科学院资深院

士、上海大学校长。中国近代力学、应用数学的奠基人之一。中国人民政治协商会议第六届、七届、八届和九届全国委员会副主席，民盟中央副主席、名誉主席。



钱伟长毕业于清华大学，加拿大多伦多大学。1956 年制定规划了中国第一次 12 年科学规划，钱伟长与钱学森、钱三强一起，被周恩来称为中国科技界的“三钱”，并被任命为清华大学副校长。钱伟长与钱学森等创办中国科学院力学研究所，同时担任中国科学院力学研究所副所长、中国科学院学术秘书、国务院科学规划委员会委员、中国科学院自动化研究所筹委会主任、波兰科学院院士。

## 二、诺贝尔物理学奖简介

**1971 年诺贝尔物理学奖**。发明并发展了全息摄影法。获奖者：D. 加博尔（英国）。评价：全息摄影法是照相术方面的成果第二次取得物理学的最高的成就。G. 李普曼的成就重现了五彩缤纷的世界。而加博尔的工作则是重现了立体图像。由于光波有两个信息，就是大小（振幅）和位相。一般的照相术只能保留大小而会丢失位相信息（所以只能得到平面像）。而加博尔通过干涉的方法可以保留两方面的信息使得光的信息完整保留。不过由于它过于精密对设备要求很高（比如平稳度非常高的平面）所以一般人很难看到全息图。全息图确实已进入我们生活的方方面面。比如全息防伪。就是你在不同距离看会出现不同的图案的防伪商标。而且全息技术更具诱惑的是可以通过改变干涉条纹（就是全息的底片）来得到不存在的图像。所以相较于全息术，PS 真是没啥技术含量。

**1972 年诺贝尔物理学奖**。从理论上解释了超导现象。获奖者：J. 巴丁、L. N. 库柏、J. R. 施里弗（美国）。评价：1911 年荷兰科学家翁纳斯（H. Kamerlingh Onnes，1853—1926）在测量低温下水银电阻率的时候发现，当温度降到－269℃附近，水银的电阻竟然消失了，从而发现了超导现象。后来人们虽然发现了很多其他物质也有超导的性质但对此的理论并不十分清楚。而这三位科学家提出的所谓 BCS（就是三人的名字）理论可以解释很大一批超导现象（主要是金属超导体）。此外值得一提的是巴丁曾在 1956 年因在晶体三极管的成就获得过诺贝尔物理学奖。成为第一个在同一领域两获诺贝尔奖的科学家（鲍林和居里是在不同领域）。

**1973 年诺贝尔物理学奖**。通过实验发现半导体中的“隧道效应”和超导物质。获奖者：江崎玲於奈（日本）、I. 贾埃弗（美国）。发现超导电流通过隧道阻挡层的约瑟夫森效应。获奖者：B. D. 约瑟夫森（英国人）。评价：隧道击穿是二极管（其实应该说 PN 结比较准确）反向被击穿的二种方式。虽然人们理论（根据量子力学，一个粒子是有可能穿过能量比它高的势垒）上很早预言了它的存在，但是却没有观察到它。而江崎玲於奈从半导体中意外观察到了它的存在。加埃沃长期致力于隧道效应的研究。1960 年于超导中发现了此现象。二者分别为隧道二极管和超导理论提供了理论依据，而当约瑟夫森年仅 20 多岁的时候就在 J. 巴丁、L. N. 库柏、J. R. 施里弗的基础上发展了超导理论为后来的超导的发展提供了基础。

**1974 年诺贝尔物理学奖**。从事射电天文学方面的开拓性研究。获奖者：M. 赖尔、A. 赫威斯（英国）。评价：天文学无疑是 20 世纪中叶最为流行的学科。随着美苏冷战，空间技术的进展一日千里。而 60 年代天文学更是出现了井喷似的发展，涌现了 60 年代天文学四大发现——星际分子，类星体，微波背景辐射和脉冲星。而在这背后离不开射电望远镜的

发展。射电望远镜主要是通过收集恒星发出的电磁波而不是一般望远镜的可见光来观测。所以对测量(某些产生非热辐射的天体,虽然不发出可见光,但往往发出强烈的射电辐射)或有很多宇宙尘埃的形体有非常重要的作用。而且为了克服电磁波造成的分辨率下降。M.赖尔还创造性地用巧妙的方法使用合成孔径技术使得孔径增大,电磁波造成的损失得以弥补。(孔径通俗地说就是望远镜镜框的大小)至于 A.赫威斯则是利用射电望远镜发现脉冲星(四大发明之一)而获奖。

**1975 年诺贝尔物理学奖**。从事原子核内部结构方面的研究。获奖者:N.玻尔,B.R.莫特尔森(丹麦)、J.雷恩沃特(美国)。评价:此次评奖又回归纯理论物理上去。这三位的贡献是提出了"原子核集体运动模型",指出了原子核内部中所有核子都会转动和振动,进一步揭示了原子核内部结构的秘密,描绘出更接近原子核结构实际的物理图像。另外这位玻尔就是那位伟大的物理学家尼尔斯玻尔的儿子。

**1976 年诺贝尔物理学奖**。发现了后来称为 J/ψ 的新粒子。获奖者:B.里克特(美国)、丁肇中(美籍华人)。评价:粒子物理学从 1932 年正电子开始就是诺贝尔奖的常客,到了 50 年代,反质子、π 介子的陆续发现都获取了这一至高荣誉,其盖尔曼的夸克模型理论,解释了这些粒子,其预言的 Ω－粒子又被实验证实。这时粒子物理学似乎已经达到了顶峰,无事可做了。然而,正是在这一短暂的沉静时期,1974 年里克特、丁肇中分别独自发现了一种新的粒子。并把夸克的种类从 3 种增加到 4 种(现在已知有六种)。注:J/ψ 子的名字其实只是二组人发现后取名不同造成的。

**1977 年诺贝尔物理学奖**。从事磁性和无序系统电子结构的基础研究。获奖者:P.W.安德逊。J.H.范弗莱克(美国)、N.F.莫特(英国)。评价:莫特开拓了作为固体物理新领域的非晶态物质电子过程的研究,被誉为这个新的分支学科的奠基人。范弗莱克从量子力学的角度解释一些磁性因此被誉为"现代磁学之父"。安德逊建立了著名的"安德逊模型",对于解释一些物理基本问题有重大意义诸如超导临界温度(即超导体电阻突然消失的温度称为临界温度)。

**1978 年诺贝尔物理学奖**。从事低温学方面的研究。获奖者:P.卡皮察(苏联)、彭齐亚斯、R.W.威尔逊(美国)。发现宇宙微波背景辐射。评价:卡皮察发现温度低于一个值时流过狭缝的液态氦的流速与压差无关,并得出结论:温度低于那个值,液态氦是超流动的,液体内部以及液体与器壁之间都没有摩擦力。卡皮察是苏联物理学的领袖人物,是朗道的良师益友,可以说在道德和学术方面皆为楷模。朗道由于树敌太多在斯大林时期曾被陷害怀疑是德国间谍而被克格勃带走。卡皮察后来亲自给斯大林写信并且说自己在液氦方面的研究只有朗道能帮助自己,并以辞职作威胁,为朗道的最终获释提供了巨大帮助(当然还有很多其他人也出了力包括玻尔),正如朗道所言:"后来是在那些年月,卡皮察的举动需要大勇、大德和水晶般纯洁的人格。"他以后始终对卡皮察怀着感激之情,曾经这样评价卡皮察:"他拥有一个科学家可能向往的一切:他的著作得到首肯,他有才华横溢的门生……但卡皮察依然孜孜不倦地从事科学研究,他的好奇心和创造力依然无穷无尽……"。

**1979 年诺贝尔物理学奖**。在发展基本粒子之间的弱点相互作用理论的贡献,特别是预言了弱中性流。获奖者:S.L.格拉肖、S.温伯格(美国)、A.萨拉姆(巴基斯坦)。评价:科学家把这个世界划分为四种力,强相互作用力、弱相互作用力、电磁力和引力,而科学家一直将这四种理论的合而为一建立统一场理论视作终极目标。而此三人的工作就是把弱相互作用力,电磁力统一起来为统一理论迈出了关键一步。这项成就被誉为 20 世纪物理学的巅峰

成就。更值得一提的是这里面也有杨振宁的功劳，其杨-米尔斯理论是推导这个理论的关键一步。可以说杨振宁也无愧为 20 世纪最伟大的物理学家之一。

威尔逊发现的微波背景辐射是天文学 60 年代四大发现之一，说明宇宙本身并不是一个安静的体系。他会通过宇宙本身发出的电磁波，应该是“大爆炸宇宙起源会留下射电噪声残余物”为大爆炸学说提供依据。大爆炸学说是关于宇宙起源最为流行的说法，简单的说即宇宙是从一个无限小的一个点经过一次大爆炸形成了我们今天的世界，预言存在弱中性流，并对基本粒子之间的弱作用和电磁作用的统一理论做出贡献。

**1980 年诺贝尔物理学奖**。发现中性 K 介子衰变中的宇称电荷(CP)不守恒。获奖者：J. W. 克罗宁、V. L. 菲奇(美国)。评价：杨振宁、李政道提出宇称不守恒是对物理学美感的极大破坏。不过人们后来发现在弱相互作用下还是满足宇称、电荷联合守恒。但是事实却是残酷的。人们发现在某种情况下 CP 守恒也是不成立的。似乎上帝所创造的世界不是那么简单能被了解的。基本粒子理论也从宇称(P)守恒，进到普遍的 CP(即用它的反电荷在取它的镜像遵循的物理定律不变)守恒，再进到更加普遍的 CPT(基本粒子的电荷，宇称两个性质都换成相反的，并倒转时间箭头，所得到的粒子将与原来的粒子遵循同样的物理规律)守恒，标志着它们的普遍性程度的不断提高也标志着人们对物质的理解更加深入。注：因为一个亚原子事件看起来既可以是在时间中向前推进，也可以是在时间中向后倒退。

## 思考与练习

1. 惠更斯原理与惠更斯-菲涅耳原理有什么不同？
2. 费马原理与几何光学等光程性的联系是什么？
3. 巴比涅原理的物理本质是什么？
4. 为什么说光的衍射与几何光学仪器的分辨本领息息相关？
5. 什么是色光三基色原理(三原色原理)？什么是颜料三原色原理？
6. 三基色原理(三原色原理)与三补色原理的区别是什么？
7. 推导菲涅耳公式并讨论其物理意义。
8. 全息照相与光的干涉和衍射的关系是什么？
9. 光波的半波损失与机械波的半波损失有何不同？
10. 布儒斯特定律有哪些应用？

# 第10章

# 光的波粒二象性与光速

关于光的本性问题,是科学家们一直探寻的课题,也是物理学发展中,特别是量子力学发展中的一个重要问题,曾有过多种观点。在17世纪末,牛顿提出光的粒子说,指出光是一种粒子,这种观点能很好地解释光的反射、折射等现象。从19世纪初到19世纪中叶,托马斯·杨、菲涅耳和麦克斯韦等一大批物理学家综合实验和理论两个方面的成果,建立了光的电磁理论,指出光是一种电磁波,即光的波动说,从而解释了光的干涉、衍射、偏振和色散等现象。

一个理论若被实验所证实,则说明它具有一定的正确性,光的粒子假说被光的反射、折射、黑体辐射、光电效应和康普顿效应等实验所证实,说明它是正确的;而光的波动说也被大量干涉、衍射和偏振实验所证实,这是客观事实。因此,在对光的本性的解释上,不应在光的粒子论与波动论之间进行取舍,应该把它们看作是光的本性的不同侧面的反映,德布罗意正是运用对称思维和类比思维的方法,大胆提出了物质波的思想——实物粒子也具有波动性,从而揭示了在光的传播过程中表现出波的特性,而在与物质相互作用的过程中,表现出粒子的特性。就是说,光具有波和粒子这两方面的特性,即光的波粒二象性。

由理论和实验所得到的结果表明,无论是静止质量为零的光子,还是静止质量不为零的电子、质子、原子等实物粒子,都同时具有波动性和粒子性,也就是具有波粒二象性。波粒二象性是现代物理学的基石之一。

## 10.1 光的粒子性的三个典型实验

光的干涉和衍射现象以及光的电磁理论从实验和理论两方面充分肯定了光的波动性,用光的电磁波理论可以圆满地解决光在传播过程中几乎所有的波动特征,但是20世纪初发现的黑体辐射、光电效应等现象却显然偏离了光的波动学说。提示把光看作波动的局限性。以下讨论光的粒子性的三个典型实验。

### 10.1.1 黑体辐射

**1. 热辐射、基尔霍夫定律**

按照物理学的观点,任何物体在不是绝对零度的任何温度下都要发射各种波长的电磁波,并且其辐射能量的大小及辐射能量按波长的分布都与温度有关,由于物质中的分子、原子受到热激发而发射电磁波的现象称为热辐射。

实验观察太阳和白炽灯中发射的光,可以发现热辐射的光谱是连续的,并且辐射谱的性质与温度有关。大多数物体辐射不可见的红外光,但当物体被加热到500℃左右时,开始发出暗红色的可见光,随着温度的不断上升,辉光逐渐亮起来,而且波长较短的辐射越来越多,大约在1500℃时就变成明亮的白炽光,这说明同一物体在一定温度下所辐射的能量,在不

同光谱区域的分布是不同的，而且温度越高，光谱中与能量最大的辐射相对应的频率也越高。此外，实验中还发现，在一定的温度下，不同物体所辐射的光谱成分有显著不同。例如，将钢加热到 800℃左右时，就可观察到明亮的红色光，但在同一温度下，溶化的水晶却不辐射可见光。我们知道，热辐射不一定需要高温，任何温度的物体都发出一定的热辐射。物体会辐射能量，同样也会吸收能量，物体在某个频率范围内发射电磁波能力越大，则它吸收该频率范围内电磁波的能力也越大。辐射和吸收的能量恰相等时称为热平衡，此时温度恒定不变。实验表明：物体辐射能多少决定于物体的温度、辐射的波长、时间的长短和发射的面积。

从实验结果可知，在单位时间内从物体单位面积向各个方向所发射的、频率在 $\nu$ 到 $\nu+\mathrm{d}\nu$ 范围内的辐射能 $\mathrm{d}W$ 与 $\nu$ 和 $T$ 有关，而且当 $\mathrm{d}\nu$ 取得足够小时，可认为辐射能 $\mathrm{d}W$ 与 $\mathrm{d}\nu$ 成正比，即

$$\mathrm{d}W=M(\nu,T)\mathrm{d}\nu \tag{10-1-1}$$

式(10-1-1)中 $M(\nu,T)$是 $\nu$ 和 $T$ 的函数，叫作该物体在温度 $T$ 时发射频率为 $\nu$ 的单色出射度(简称单色辐出度)。

从物体表面单位面积上发出的各种频率的总辐射功率称为物体的辐射出射度(简称辐出度)，用 $M_0(T)$表示，在一定温度 $T$ 下，这个量

$$M_0(T)=\int_0^\infty \mathrm{d}W=\int_0^\infty M(\nu,T)\mathrm{d}\nu \tag{10-1-2}$$

显然，$M_0(T)$只是温度的函数，它对于各种不同的物体，特别在表面的情况不同时，单色辐出度 $M(\nu,T)$是不同的，相应地 $M_0$ 也是不同的。

另一方面，当辐射照射到某一不透明物体表面时，其中一部分能量被物体散射或者反射，另一部分能量则被物体吸收，如以 $\mathrm{d}W$ 表示频率在 $\nu$ 到 $\nu+\Delta\nu$ 范围内，照射到温度为 $T$ 的物体的单位面积上的辐射能，$\mathrm{d}W'$表示物体单位面积上所吸收的辐射能量，那么这二者的比值

$$A(\nu,T)=\mathrm{d}W'/\mathrm{d}W \tag{10-1-3}$$

叫作该物体的吸收比，按定义吸收比总是满足 $0\leqslant A(\nu,T)\leqslant 1$，并且一切物体的吸收比都是随物体的温度和入射的辐射频率而改变的，不同物体特别是各种不同的表面，吸收比也是不同的。

$N$ 个不同的物体置于一绝热恒温体内，经过热辐射交换能量，达到热平衡态。要维持温度不变，则物体吸收的辐射能必须等于辐射出去的能量。将温度不同的物体 $A_1$、$A_2$ 和 $A_3$ 放到一个密闭的理想绝热容器里，如图 10-1-1 所示，如果容器内部是真空的，则物体与容器之间以及物体之间只能辐射或吸收来交换能量。如果单位时间内辐射体发出的能量比吸收的能量多时，它的温度下降，这时辐射就会减弱，反之，辐射体的温度将升高，所有物体包括容器在内都会达到相同的温度，建立热平衡，此时各物体在单位时间内发出的能量恰好等于吸收的能量。由此可见，在热平衡的情况下，单色辐出度较大的物体，其吸收比也一定较大；单色辐出度较小的物体，其吸收比也一定较小。1859 年，基尔霍夫根据热平衡原理得出这样一个定律：物体的单色辐出度 $M(\nu,T)$和吸收比 $A(\nu,T)$的比值与物体的性质无关，而只是频率和温度的普适函数



图 10-1-1　热辐射交换能量

$$\frac{M(\nu,T)}{A(\nu,T)}=f(\nu,T) \tag{10-1-4}$$

**2. 黑体与黑体辐射**

绝对黑体(简称黑体)是指一个能够全部吸收投射在它上面的辐射而无反射的物体。因为所有的物体都会发射和吸收电磁波,所以,绝对黑体是不存在的。但是,要在实验中制作一个简单的近似的黑体却并不难。如图 10-1-2 所示,可以在封闭的空腔壁上开一个小孔,外来的电磁波由小孔进入空腔,由于电磁波再次从小孔穿出来的概率近乎为零,也就是说带小孔的空腔几乎可吸收入射于其中的全部辐射,所以,可以把它近似看作黑体。

当空腔壁与腔内的辐射处于热平衡时,腔内单位面积所发出的辐射能量就和它所吸收的辐射能量相等。实验得知,平衡时辐射能量密度按波长分布的曲线,其形状和位置只与黑体的绝对温度有关,而与空腔的形状和组成物质无关。所谓黑体辐射其实就是物质达到热平衡时,黑体具有最大的吸收比,因而它也具有最大的单色辐出度。黑体的吸收比与频率和温度无关,它是等于 1 的常数,设以 $M_b(\nu,T)$ 和 $A_b(\nu,T)$ 表示黑体的单色辐出度和吸收比,由于 $A_b(\nu,T)=1$,则基尔霍夫定律可写成

$$\frac{M_b(\nu,T)}{A_b(\nu,T)}=M_b(\nu,T)=f(\nu,T) \tag{10-1-5}$$

19 世纪末,物理学家从实验和理论两方面都研究了各种温度下的黑体辐射,测量了它们的单色辐出度按波长分布的情况,得出如图 10-1-3 的实验曲线,可以看出:每一条曲线都有一个极大值,随着温度的升高,黑体的单色辐出度迅速增大,并且曲线的极大值逐渐向短波方向移动。

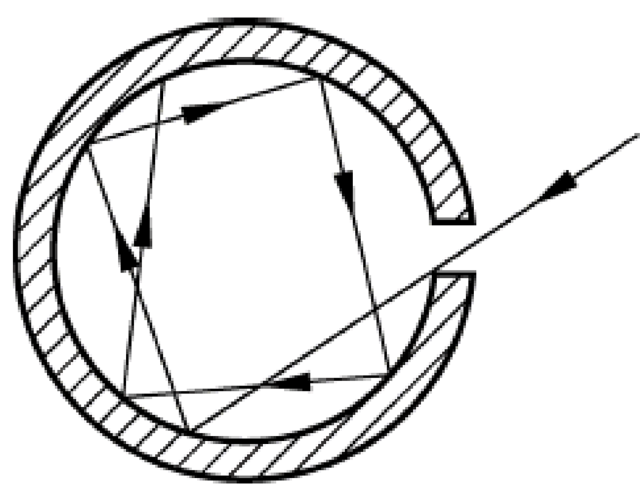


图 10-1-2 空腔黑体辐射

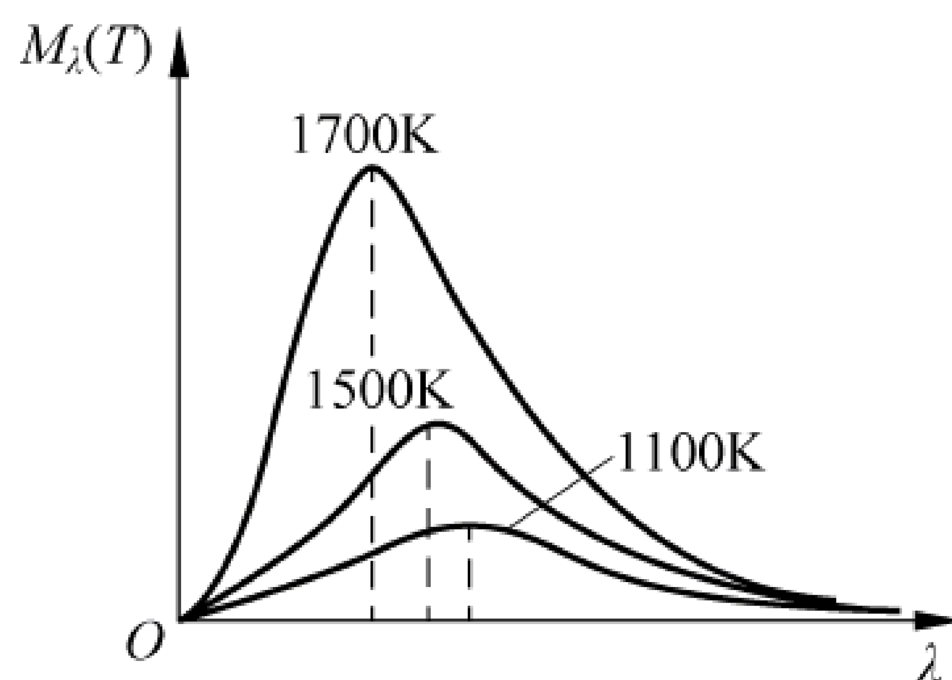



图 10-1-3 黑体辐射单色辐出度

我们从式(10-1-5)可以看出,黑体的单色辐出度 $M_b(\nu,T)$ 等于普适函数 $f(\nu,T)$,因此研究 $M_b(\nu,T)$ 就成为研究黑体辐射乃至热辐射的关键,找出这个普适函数 $f(\nu,T)$ 的形式,也就是从理论上解释实验所得的黑体辐射能量的分布曲线,这是热辐射理论的基本问题。而这个问题的解决,在历史上是有一个发展过程的。

1879 年斯特藩(Josef Stefan,1835—1893)在实验中发现黑体的辐出度 $M(T)$ 与黑体的热力学温度 $T$ 的四次方成正比,1884 年,玻耳兹曼从理论上得出

$$M_0(T)=\int_0^{\infty} M_b(\nu,T)\mathrm{d}\nu=\sigma T^4 \tag{10-1-6}$$

其中,$\sigma=5.67051\times10^{-8}\,\mathrm{W/(m^2K^4)}$ 是一个普适常量,叫作斯特藩-玻耳兹曼常量。

当黑体的热力学温度升高时,与单色辐出度 $M_b(\nu,T)$ 的峰值相对应的波长向短波方向移动,经理论推导得出

$$M_b(\nu,T)=c\nu^3 f\left(\frac{\nu}{T}\right) \tag{10-1-7}$$

$$T\lambda_m=b \tag{10-1-8}$$

式(10-1-8)即是维恩(Wilhelm Carl Werner Otto Fritz Franz Wien,1864—1928)位移定律,维恩为此获得诺贝尔物理学奖。其中 $b=2.8978\times10^{3}\,\mathrm{mK}$,是一个和温度无关的量。

图 10-1-4 说明了在某温度下的实验曲线与维恩公式符合的情况：维恩公式在短波部分与实验结果非常符合,但是在长波部分则显著不一致。

1900 年,瑞利与金斯(J. H. Jeans,1877—1946)按照经典电磁场理论和经典统计物理进行计算波长从 $l$ 到 $l+\mathrm{d}l$ 的辐射能量,得到如下公式：

$$M_{\lambda}(T)\mathrm{d}\lambda=\frac{2\pi c}{\lambda^{4}}kT\,\mathrm{d}\lambda \tag{10-1-9}$$

$$M_{\nu}(T)\mathrm{d}\nu=\frac{2\pi\nu^{2}}{c^{2}}kT\,\mathrm{d}\nu \tag{10-1-10}$$

图 10-1-5 表示这个函数在长波区域和实验曲线很好地符合,但随着频率的增大,与实验的差距越来越大。

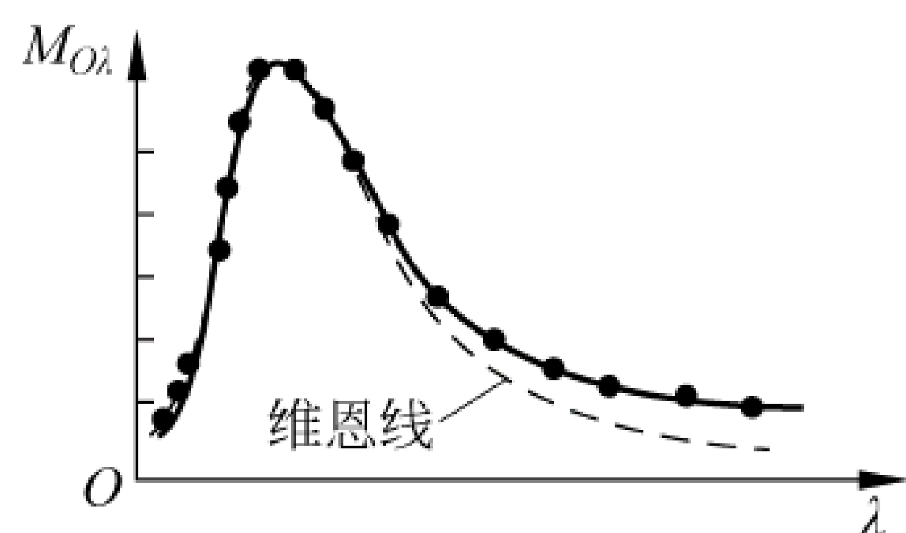



图 10-1-4　维恩位移定律

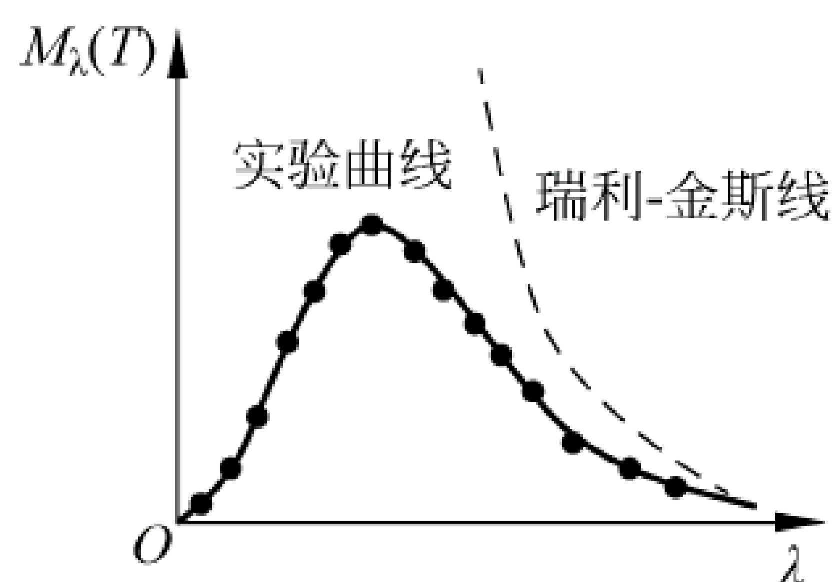



图 10-1-5　瑞利与金斯定律

1900 年,普朗克为了得到与实验曲线相一致的公式,提出了与经典物理学概念截然不同的新假设。

① 金属空腔壁中电子的振动可视为一维谐振子,它吸收或发射电磁辐射能量时,以与振子的频率成正比的能量子 $\varepsilon=h\nu$ 为基本单元来吸收或发射能量。

② 空腔壁上带电谐振子所吸收或发射的能量是 $h\nu$ 的整数倍。

$$E=nh\nu,\quad n=1,2,\cdots \tag{10-1-11}$$

其中,普朗克常数 $h=6.626\times10^{-34}\,\mathrm{Js}$。普朗克当时不得不假定：当带电谐振子与周围电磁场交换能量时,放出能量不能以连续的方式,只能以量子方式进行。

普朗克经理论推导,得出温度为 $T$ 的黑体,在单位面积上频率在 $\nu+\mathrm{d}\nu$ 范围内辐射的能量为

$$M_{\nu}(T)\mathrm{d}\nu=\frac{2\pi h\nu^{3}}{c^{2}}\frac{\mathrm{d}\nu}{\mathrm{e}^{h\nu/kT}-1} \tag{10-1-12}$$

这就是普朗克黑体辐射公式。这个公式不仅与实验结果相符合,而且也解决了在经典热力学中固体比热与实验不符的问题。

当波长较大时,即 $kT\gg h\nu$,瑞利-金斯公式是普朗克公式的极限结果。当波长较短时,即 $kT\ll h\nu$,即

$$\mathrm{e}^{h\nu/kT}-1=\left(1+\frac{h\nu}{kT}+\cdots\right)-1\approx\frac{h\nu}{kT} \tag{10-1-13}$$

代入普朗克公式即得

$$M_{\nu}(T)\mathrm{d}\nu=\frac{2\pi\nu^{2}}{c^{2}}kT\,\mathrm{d}\nu \tag{10-1-14}$$

我们还可以从普朗克公式导出斯特藩定律和维恩定律。

## 10.1.2 光电效应

### 1. 光电效应的发现

1886—1887年,德国物理学家赫兹在证明电磁波的存在和光的麦克斯韦电磁波理论的过程中,发现当两电极之一受紫外光照射时,两电极容易发生放电现象,光竟奇妙地导致了电。直到电子发现以后,人们才知道此现象是电子在光照作用下从金属表面逃逸而出造成的。实验装置如图10-1-6所示,由光电管、微电流测试仪、光源及滤色片四部分构成。

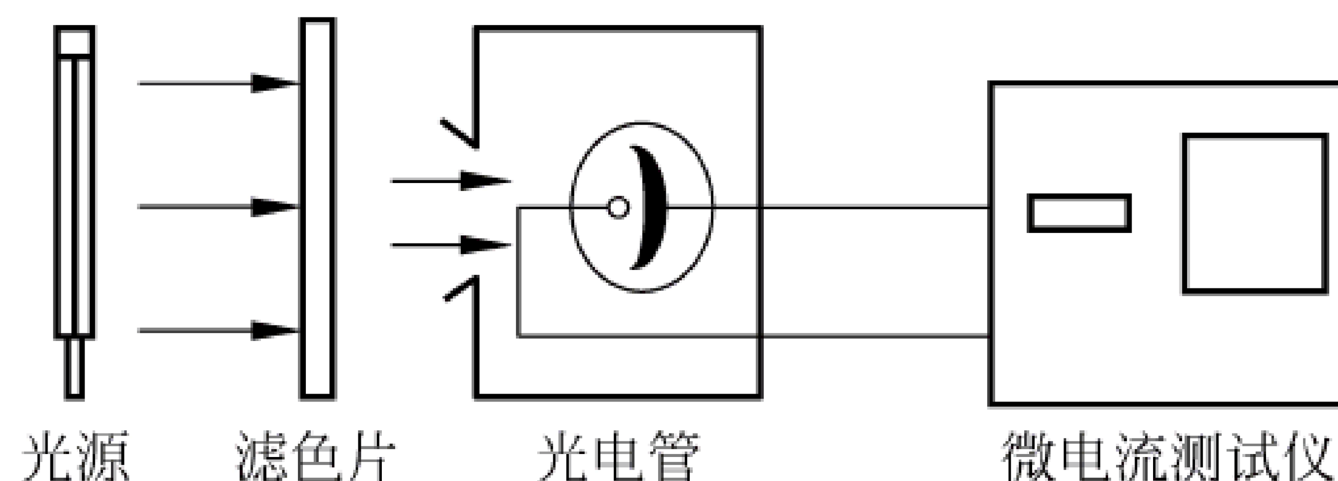



图10-1-6　光电效应实验装置

这种电子在光的作用下从金属表面发射出来的现象就称为光电效应,逸出来的电子就称为光电子。光电效应得到了一些与传统理论截然不同的实验结果:光电流的大小与入射光的强度成正比;光电子的能量与光的频率大小有关;光电效应具有频率红限,与时间的积累无关。这些都是经典的物理理论所无法解释的。直到发现电子后,人们才知道这是由于紫外光的照射使大量电子从金属表面逸出的缘故。

### 2. 光电效应实验规律

(1) 饱和电流 $I_m$ 的大小与入射光的强度成正比,也就是单位时间内逸出的光电子数目与入射光的强度成正比;

(2) 光电子的最大初动能(或遏止电压)与入射光的强度无关,而只与入射光的频率有关,频率越高,光电子的能量就越大,即所谓的频率"红限";

(3) 光的照射和光电子的逸出几乎是同时的,在测量的精度范围内($<10^{-9}$ s)观察不出这两者间存在滞后现象;

(4) 入射光的频率低于"红限",无论光的强度多大,照射时间多长,都不能使光电子逸出。

### 3. 爱因斯坦的光电效应量子观

为了尽量缩小与经典物理学之间的差距,普朗克把能量子的概念与振子辐射能量的过程相互关系,而认为辐射场本身仍然是连续的电磁波。普朗克提出了能量子概念之后,许多物理学家都想从经典物理学中求得对光电效应的解释,但始终无法成功。直到1905年,爱因斯坦在光电效应的研究中,才突破了普朗克的认识,推广了普朗克的能量子 $\varepsilon_0=h\nu$ 的概念,提出了光子与光电子观点,认为光的能量单元是光子:$E=h\nu$。电子对光子的吸收主要表现为单光子的吸收(双光子及多光子吸收的概率极小),于是,爱因斯坦提出了运用能量单元是光子的思想来解释光电效应现象,其方程为

$$h\nu=\frac{1}{2}mv^2+W \tag{10-1-15}$$

其中,$h$ 为普朗克常量,$\nu$ 为光的频率,$\frac{1}{2}mv^2$ 为光电子逸出的初动能,$W$ 为金属的逸出功。

爱因斯坦光电效应量子观成功地解释了光电效应的现象和规律:

第一,对于某种确定的金属,逸出功为一常量。如果入射光的频率过低,即 $h\nu<W$,而一个电子只能吸收一个光子,所以电子就根本不能脱离金属表面,这时即使入射光很强(光

子数很多），仍不会产生光电效应；

第二，由光电效应方程可知，对于给定的金属，光子的频率越高，光子的能量就越大，电子吸收光子后的光电子的动能就越大，这与入射光的强度（光子个数的多少）无关；

第三，金属中的电子能一次全部吸收入射的一个光子，因此，光电效应的产生就不需要积累能量的过程，即光电效应发生的时间极短。

对于光电效应，我们还可以进一步理解，$\frac{1}{2}mv^2$ 称为光电子的最大初动能，原因在于 $W$ 是光电子逸出金属表面所需的最小能量。可以这样认为，逸出功是指最表层的电子脱离金属时挣脱金属对它束缚的能量。其实，在光电效应中产生的光电子并非都是金属最表面层的电子，这些光电子的初动能显然小于最表层逸出的光电子。另外，有一部分电子被大于该金属的极限频率的光照射后向金属内部运动，这些电子不会变成光电子。所以，某种金属的极限频率应理解为恰好使金属表层电子离开金属时光的频率。

### 10.1.3　康普顿效应

**1. 康普顿效应实验结果**

光的量子的、微粒的性质，尤其是光子具有能量、质量、动量以及光在和物质发生作用时上述量的守恒性，在康普顿（Arthur Holly Compton，1892—1962）于1922年观察的X射线散射实验现象中，更明显地表现出来。由于X射线的波长很短，所以，即使通过不含杂质晶体中的这种散射时，发现散射谱线中除了波长和原射线相同的成分以外，还有一些波长较长的成分，两者的波长差随着散射角的大小而变，这种波长改变的散射称为康普顿效应。从波动理论来看，散射光是由电子受到入射光的作用作受迫振动并向各个方向发出的次波所引起的，按照经典电磁波理论，受迫振动的频率和散射光的频率都应与入射光的频率相同，显然散射光波长发生改变的康普顿效应又是难以用波动理论观点来解释的。

康普顿提出，将 $\lambda=0.07078\text{nm}$ 的X光投射到石墨上，然后在不同的角度测量被石墨分子散射的X光强度，实验装置如图10-1-7所示。当 $\theta=0$ 时，散射光只有等于入射频率的单一频率光。当 $\theta\neq0$（如45°、90°、135°）时，发现存在两种频率的散射光，如图10-1-8所示。一种频率与入射光相同，另一种频率比入射光低。后者随角度增加偏离增大。

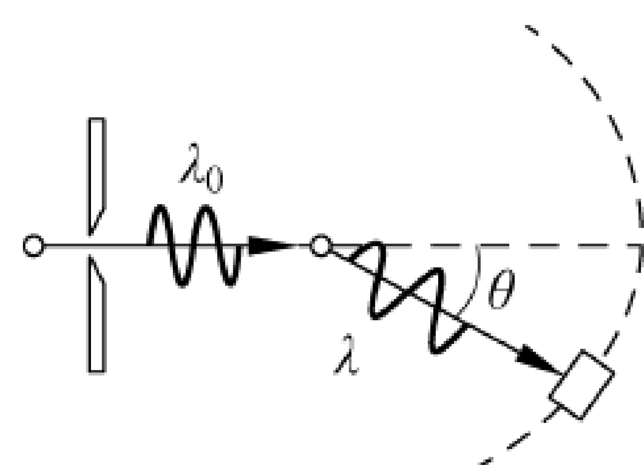


图10-1-7　康普顿效应实验装置

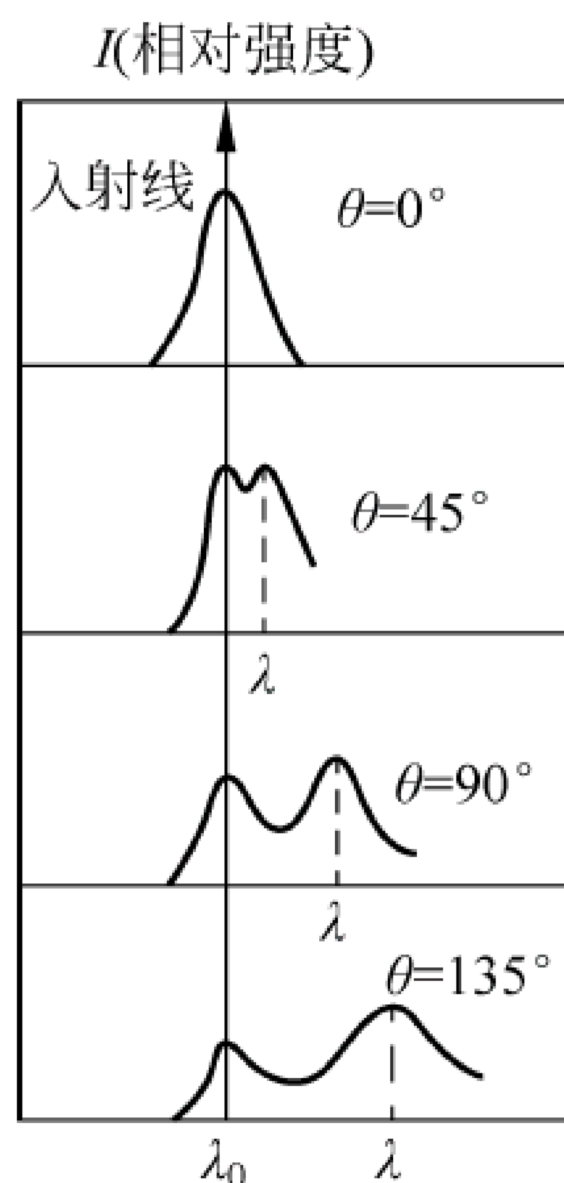


图10-1-8　康普顿散射光

**2. 康普顿效应的解释**

(1) 经典解释(电磁波的解释)

单色电磁波作用于比波长尺寸小的带电粒子上时,引起受迫振动,向各方向辐射同频率的电磁波。经典理论解释频率不变的一般散射可以,但对康普顿效应不能做出合理解释!

(2) 光子理论解释

X射线为一些能量特别大的 $\varepsilon=h\nu$ 的光子,与自由电子发生完全弹性碰撞,电子获得一部分能量,散射的光子能量减小,频率减小,波长变长。如图10-1-9所示。设这个过程的动量守恒与能量守恒仍成立,则由电子与X射线光子构成的系统动量守恒与能量守恒即可得

$$\lambda-\lambda_0=\frac{2h}{m_0c}\sin^2\frac{\theta}{2} \qquad (10\text{-}1\text{-}16)$$

其中 $\frac{h}{m_0c}=0.00241\text{nm}$,称为康普顿波长。

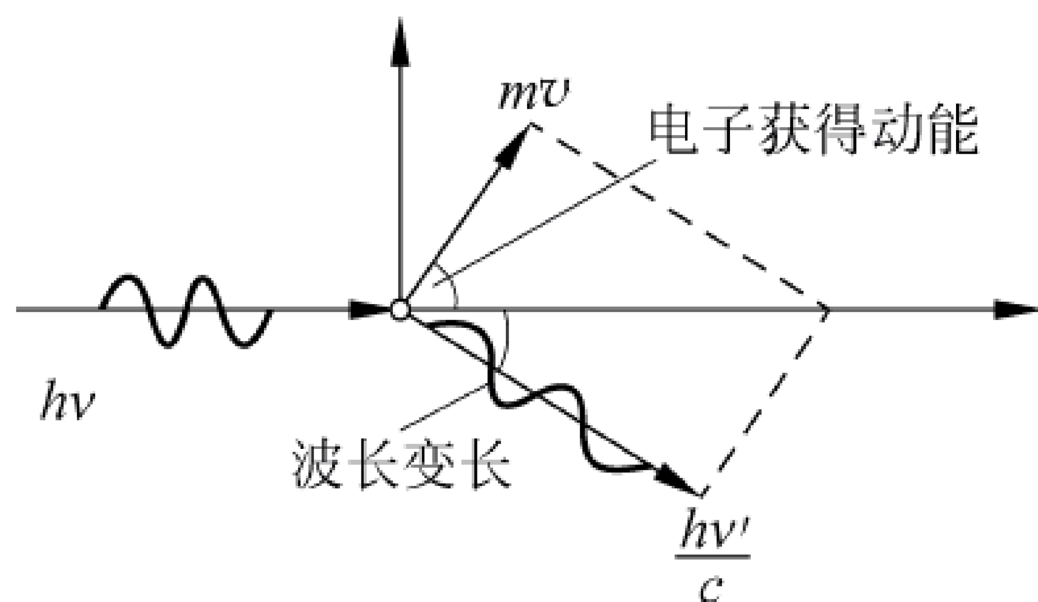



图10-1-9 康普顿散射光

**3. 注意几点**

(1) 散射波长改变量的数量级为 $10^{-12}$ m,对于可见光波长数量级约为 $10^{-7}$ m,远小于此,所以,观察不到康普顿效应。

(2) 散射光中有与入射光相同的波长的光线,是由于其光子与原子碰撞,原子质量很大,与光子碰撞后,原子能量不变,其散射光频率不变。

康普顿效应的发现,以及理论分析和实验结果的一致,不仅有力地证实了光子假说的正确性,并且证实了微观粒子的相互作用过程中,也严格遵守能量守恒和动量守恒定律。

## 10.2 关于光的波粒二象性中的辩证思维

自从托马斯·杨在1801年实现了普通光源的双光束干涉以后,经过菲涅耳等人大量的理论和实验研究,人们对光的波动本性深信不疑,光的干涉、衍射和偏振现象都可以用光的波动理论给以满意的解释,而且从理论和实验上都证明了光的直线传播是光的衍射的极限情况。然而还有许多与光有关的实验现象是用波动理论无法解释的。热辐射、光电效应和康普顿散射、原子线状光谱以及原子的稳定性等问题揭示了光的另一个方面的特性,即光的粒子性,从而使人们对光有了进一步的认识,即光具有波粒二象性。

### 10.2.1 光的波动说与光的粒子说的发展史

早先关于光的本性的概念,是以光的直线传播为基础的,但从17世纪开始,就发现了与光的直线传播不完全符合的事实。19世纪中期,意大利学者格里马第(Francesco Maria Grimaldi,1618—1663)设计了一个实验:让一束光穿过一个小孔,让这束光穿过小孔后照到暗室里的一个屏幕上。他发现光线通过小孔后的光影明显变宽了。格里马第进行了进一步的实验,他让一束光穿过两个小孔后照到暗室里的屏幕上,这时得到了有明暗条纹的图像。他认为这种现象与水波十分相像,从而得出结论:光是一种能够作波浪式运动的流体,光的不同颜色是波动频率不同的结果。格里马第第一个提出了“光的衍射”这一概念,是光的波动学说最早的倡导者。1663年,英国科学家波意耳(Robert Boyle,1627—1691)提出了物体的颜色不是物体本身的性质,而是光照射在物体上产生的效果。他第一次记载了肥皂泡和玻璃球中的彩色条纹。这一发现与格里马第的说法有不谋而合之处,为后来的研究奠

定了基础。不久后，英国物理学家胡克(Robert Hooke，1635—1703)重复了格里马第的实验，并通过对肥皂泡膜的颜色的观察提出了“光是以太的一种纵向波”的假说。根据这一假说，胡克也认为光的颜色是由其频率决定的。

1672年，牛顿在他的论文《关于光和色的新理论》中谈到了他所作的光的色散实验：让太阳光通过一个小孔后照在暗室里的棱镜上，在对面的墙壁上会得到一个彩色光谱。他认为，光的复合和分解就像不同颜色的微粒混合在一起又被分开一样。在这篇论文里他用微粒说阐述了光的颜色理论。第一次波动说与粒子说的争论由“光的颜色”这根导火索引燃了。从此胡克与牛顿之间展开了漫长而激烈的争论。

波动说的支持者，荷兰著名天文学家、物理学家和数学家惠更斯继承并完善了胡克的观点。惠更斯早年在天文学、物理学和技术科学等领域做出了重要贡献，并系统地对几何光学进行过研究。1666年，惠更斯应邀来到巴黎科学院以后，并开始了对物理光学的研究。在他担任院士期间，惠更斯曾去英国旅行，并在剑桥会见了牛顿。二人彼此十分欣赏，而且交流了对光的本性的看法，但此时惠更斯的观点更倾向于波动说，因此他和牛顿之间产生了分歧。正是这种分歧激发了惠更斯对物理光学的强烈热情。回到巴黎之后，惠更斯重复了牛顿的光学试验。他仔细地研究了牛顿的光学试验和格里马第的实验，认为其中有很多现象都是微粒说所无法解释的。因此，他提出了波动学说比较完整的理论。他认为用光的粒子说解释牛顿环的问题的时候存在问题，同时光的粒子说也难以说明绕过障碍物之后所形成的衍射现象。

惠更斯认为，光是一种机械波；光波是一种靠物质载体来传播的纵向波，传播它的物质载体是“以太”；波面上的各点本身就是引起媒质振动的波源。根据这一理论，惠更斯证明了光的反射定律和折射定律，也比较好地解释了光的衍射、双折射现象和著名的“牛顿环”实验。如果说这些理论不易理解，惠更斯又举出了一个生活中的例子来反驳微粒说。如果光是由粒子组成的，那么在光的传播过程中各粒子必然互相碰撞，这样一定会导致光的传播方向的改变。而事实并非如此。

就在惠更斯积极地宣传波动学说的同时，牛顿的微粒学说也逐步建立起来了。牛顿修改和完善了他的著作《光学》。基于各类实验，在《光学》一书中，牛顿一方面提出了两点反驳惠更斯的理由：第一，光如果是一种波，它应该同声波一样可以绕过障碍物、不会产生影子；第二，冰洲石的双折射现象说明光在不同的边上有不同的性质，波动说无法解释其原因。另一方面，牛顿把他的物质微粒观推广到了整个自然界，并与他的质点力学体系融为一体，为微粒说找到了坚强的后盾。所以，在光的粒子说与波动说的争论时期，牛顿的粒子说占据了统治地位。当时，大部分的物理学家都支持粒子说。

18世纪末，在德国自然哲学思潮的影响下，人们的思想逐渐解放。英国著名物理学家托马斯·杨开始对牛顿的光学理论产生了怀疑。根据一些实验事实，杨氏于1800年写成了论文《关于光和声的实验和问题》。在这篇论文中，杨氏把光和声进行类比，因为二者在重叠后都有加强或减弱的现象，他认为光是在以太流中传播的弹性振动，并指出光是以纵波形式传播的。他同时指出光的不同颜色和声的不同频率是相似的。

1801年，杨氏进行了著名的杨氏双缝干涉实验。实验所使用的白屏上明暗相间的黑白条纹证明了光的干涉现象，从而证明了光是一种波。同年，杨氏在英国皇家学会的《哲学会刊》上发表论文，分别对“牛顿环”实验和自己的实验进行解释，首次提出了光的干涉的概念和光的干涉定律。

1803年,杨氏写成了论文《物理光学的实验和计算》。他根据光的干涉定律对光的衍射现象作了进一步的解释,认为衍射是由直射光束与反射光束干涉形成的。但由于他认为光是一种纵波,所以在理论上遇到了很多麻烦。他的理论受到了英国政治家布鲁厄姆尖刻的批评,被称作是"不合逻辑的""荒谬的""毫无价值的"。

虽然杨氏的理论以及后来的辩驳都没有得到足够的重视,甚至遭人毁谤,但他的理论激起了牛顿学派对光学研究的兴趣。

1808年,拉普拉斯(Laplace Pierre Simon de,1749—1827)用微粒说分析了光的双折射线现象,批驳了杨氏的波动说。

1809年,马吕斯(Etienas Louts Mallis,1775—1812)在实验中发现了光的偏振现象。在进一步研究光的简单折射中的偏振时,他发现光在折射时是部分偏振的。因为惠更斯曾提出过光是一种纵波,而纵波不可能发生这样的偏振,这一发现成为反对波动说的有利证据。

面对这种情况,杨氏对光学再次进行了深入的研究,1817年,他放弃了惠更斯的光是一种纵波的说法,提出了光是一种横波的假说,比较成功地解释了光的偏振现象。吸收了一些牛顿派的看法之后,他又建立了新的波动说理论。杨氏把他的新看法写信告诉了牛顿派的阿喇果。

1817年,巴黎科学院悬赏征求关于光的干涉的最佳论文。土木工程师菲涅耳也卷入了波动说与微粒说之间的纷争。在1815年菲涅耳就试图复兴惠更斯的波动说,但他与杨氏没有联系,当时还不知道杨氏关于衍射的论文,他在自己的论文中提出,是各种波的互相干涉使合成波具有显著的强度。事实上他的理论与杨氏的理论正好相反。后来阿喇果告诉了他:杨氏新提出了光是一种横波的理论。从此,菲涅耳以杨氏理论为基础开始了他的研究。1819年,菲涅耳成功地完成了对由两个平面镜所产生的相干光源进行的光的干涉实验,继杨氏干涉实验之后再次证明了光的波动说。阿喇果与菲涅耳共同研究一段时间之后,转向了波动说。1819年底,在菲涅耳对光的传播方向进行定性实验之后,他与阿喇果一道建立了光波的横向传播理论。

1882年,德国天文学家夫琅禾费首次用光栅研究了光的衍射现象。在他之后,德国另一位物理学家施维尔德(1792-1871)根据新的光波学说,对光通过光栅后的衍射现象进行了成功的解释。至此,新的波动学说牢固地建立起来了。微粒说开始转向劣势。

随着光的波动学说的建立,人们开始为光波寻找载体,以太说又重新活跃起来。一些著名的科学家成为以太说的代表人物。但人们在寻找以太的过程中遇到了许多困难,于是各种假说纷纷提出,以太成为19世纪的热点之一。

菲涅耳在研究以太时发现的问题是,横向波的介质应该是一种类固体,而以太如果是一种固体,它又怎么能不干扰天体的自由运转呢。不久以后泊松也发现了一个问题:如果以太是一种类固体,在光的横向振动中必然要有纵向振动,这与新的光波学说相矛盾。

1887年,美国物理学家麦克耳孙与化学家莫雷以"以太漂流"实验否定了以太的存在。但此后仍不乏科学家坚持对以太的研究。甚至在法拉第的光的电磁说、麦克斯韦的光的电磁说提出以后,还有许多科学家潜心致力于对以太的研究。

19世纪中后期,在光的波动说与微粒说的论战中,波动说已经取得了决定性胜利。但人们在为光波寻找载体时所遇到的困难,却预示了波动说所面临的危机。

光的粒子性问题首先起源于热平衡物体的电磁波辐射,即所谓的热辐射。

1887 年，德国科学家赫兹发现光电效应，光的粒子性再一次被实验证明！

1900 年普朗克针对热辐射波动解释的困难，提出了电磁辐射的量子学说。

1905 年，爱因斯坦仔细分析了光电效应之后，在普朗克黑体辐射能量子假说的基础上，作了至关重要的发展。他指出：光不仅像普朗克所假定的那样在辐射和吸收时能量是一份一份的，具有量子性，在空间传播时也不像波动理论所认为的是连续分布，而是集中在一些叫光子的粒子上，每个光子有确定的能量。对波动理论中频率为 $\nu$ 的光波，对应的光子的能量 $\varepsilon = h\nu$，$h$ 是普朗克常数，在与物质相互作用时，光子只能作为一个整体被吸收或辐射。

1921 年，爱因斯坦因为“光电效应的量子解释”这一成就而获得了诺贝尔物理学奖。

1921 年，康普顿在实验中证明了 X 射线的粒子性。

1924 年，在光具有波粒二象性的启发下，德布罗意在他的博士学位论文中指出，既然过去认为是波动的光具有粒子性，那么，作为实物粒子(指有静质量的粒子)也应该与光子一样，也具有波动性。

1925 年玻恩(Born Max，1882—1970)为了解释波动性和粒子性之间的联系，提出的概率波概念。

1927 年，杰默尔(Jie Moer)和后来的乔治·汤姆孙(George Thomson)在实验中证明了电子束具有波的性质。同时人们也证明了氦原子射线、氢原子和氢分子射线具有波的性质。

1928 年，德布罗意的物质波假说得到了证实，这是戴维孙(Dawisson)和革末(Gorrner)以及汤姆孙(Thomson)努力的结果。

在新的事实与理论面前，光的波动说与微粒说之争以“光具有波粒二象性”而落下了帷幕。

光的波动说与微粒说之争从 17 世纪初笛卡儿提出的两点假说开始，至 20 世纪初以光的波粒二象性告终，前后共经历了三百多年的时间。牛顿、惠更斯、托马斯·杨、菲涅耳等多位著名的科学家成为这一论战双方的主辩手。正是他们的努力揭开了遮盖在“光的本质”外面那层扑朔迷离的面纱。

### 10.2.2　光的电磁说

光的干涉和衍射现象无可怀疑地证明了光是一种波，到 19 世纪中叶，光的波动说已经得到公认。但是，光是什么性质的波？难道像水波一样？像声波一样？光波的本质是什么，这个问题一直没有解决。那时候人们总是习惯于按照机械波的模型把光波看成是在某种弹性介质里传播的振动。

到了 19 世纪 60 年代，麦克斯韦预言了电磁波的存在，并且从理论上得出，电磁波在真空中的传播速度应为 $3.11\times10^{8}\,\mathrm{m/s}$，而当时实验测得的光速为 $3.15\times10^{8}\,\mathrm{m/s}$，两个数值非常接近。麦克斯韦认为这不是一种巧合，它表明光与电磁现象之间有本质的联系，由此他提出光在本质上是一种电磁波，这就是光的电磁说。1886—1888 年间，赫兹做了一系列实验，证实了电磁波的存在，并且测出了实验中的电磁波的频率和波长，从而计算出了电磁波的传播速度，发现电磁波的速度确实与光速相同，这样就证明了光的电磁说的正确性。

在电磁波中，能够作用于人的眼睛并引起视觉的，只是一个很窄的波段，通常叫作可见光，其中波长最短的是紫光，波长约为 370nm，波长最长的是红光，波长约为 750nm。波长更长的光不能引起视觉，叫作红外线。红外线的波长范围很宽，为 $750\sim10^{6}\,\mathrm{nm}$。

一切物体，包括大地、人体、农作物和车船，都在辐射红外线。物体的温度越高，它辐射的红外线越强，波长越短。在热学中所说的热辐射，指的就是红外线辐射，它是热传递的方

式之一。

波长比紫外线更短的光叫作伦琴射线,也叫 X 射线。是德国物理学家伦琴(Roentgen Wilhelm Conrad,1845—1923)在 1895 年发现的。它的穿透能力很强,能使包在黑纸里的照相底片感光。由于 X 射线穿透物质的厚度跟物质的密度有关,在工业上可以用来检查金属部件内部有没有砂眼、裂纹等缺陷,在医学上可以用来透视人体,检查体内的病变及骨骼情况。伦琴在发现 X 射线时于 1895 年 12 月 22 日在实验室为他夫人的手拍下的照片,是世界上第一张 X 射线的人手照片。

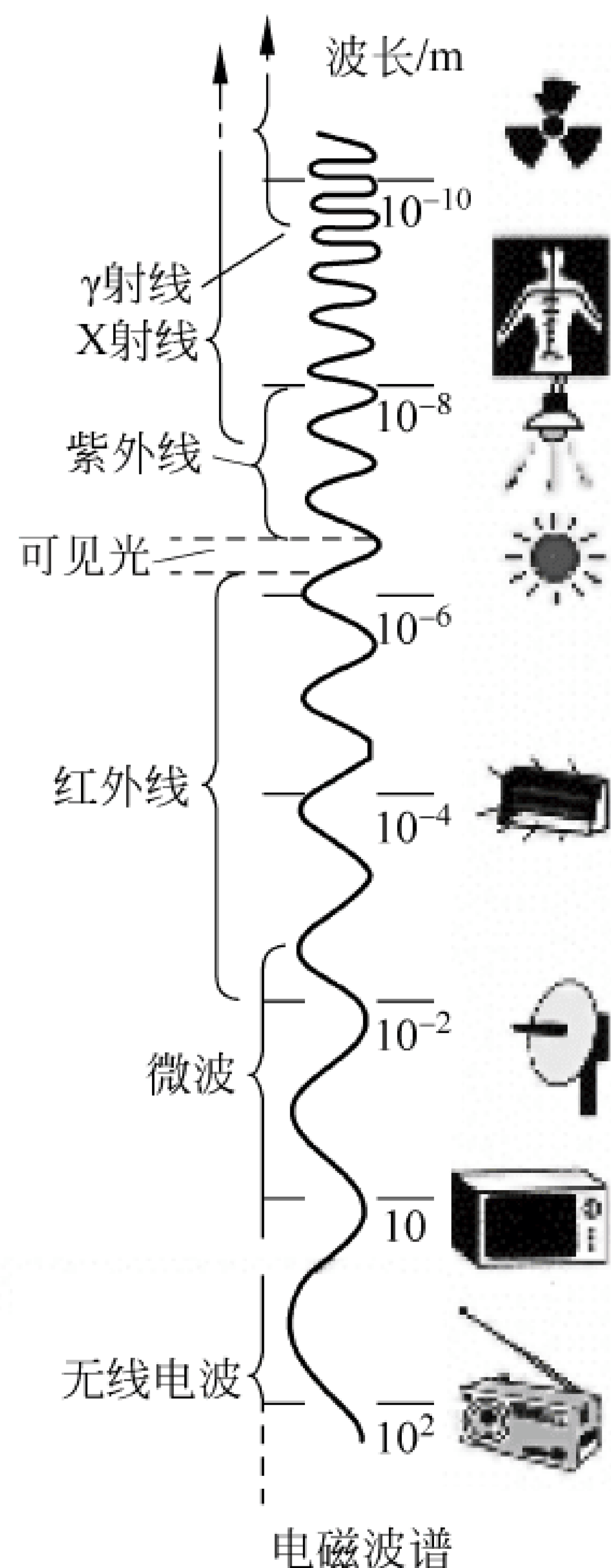


图 10-2-1 电磁波波谱

此外还有比伦琴射线波长更短的电磁波,那就是 γ 射线。无线电波、红外线、可见光、紫外线、伦琴射线、γ 射线等合起来,构成了范围非常广阔的电磁波谱,其中最长的波长是最短的波长的 $10^{21}$ 倍以上,从图 10-2-1 中我们可以看出,长波的红外线和微波已经重叠,短波的紫外线已经进入伦琴射线的区域。所以,它们之间的区别并没有绝对意义,从无线电波到 γ 射线,都是本质上相同的电磁波,它们的行为服从共同的规律。

不同的电磁波,产生的机理不同。无线电波是振荡电路中自由电子的周期性运动产生的;红外线、可见光、紫外线是原子的外层电子受到激发后产生的;伦琴射线是原子的内层电子受到激发后产生的;γ 射线是原子核受到激发后产生的。不同的电磁波,由于它们的频率或波长不同,因而表现出不同的特性。例如,波长较长的无线电波容易表现出干涉、衍射等现象,而波长越来越短的紫外线、伦琴射线和 γ 射线等,要观察到它们的干涉、衍射现象就越来越困难了。

### 10.2.3 光的粒子性

普朗克于 1900 年提出能量量子化假设,他认为当谐振子的能量状态改变时,就要一份一份地发射或吸收能量,发射或吸收的能量只是能量子的整数倍,而能量子的大小与谐振子的频率成正比。经分析知,当黑体辐射光的频率减小时,其单色输出度减小,在低频长波区,能量的不连续性可以被忽略;而当黑体辐射光频率增高时,其单色输出度增高,在高频短波区能量的不连续性比较突出,但普朗克当时仍然认为辐射场是连续的电磁波。

1905 年,年仅 26 岁的爱因斯坦仔细分析了光电效应之后,在普朗克黑体辐射能量子假说的基础上,作了至关重要的发展。他指出:光不仅像普朗克所假定的那样在辐射和吸收时能量是一份一份的,具有量子性,在空间传播时也不像波动理论所认为的是连续分布,而是集中在一些叫光子的粒子上,每个光子有确定的能量。对波动理论中频率为 $\nu$ 的光波,对应的光子的能量 $\varepsilon=h\nu$,$h$ 是普朗克常数,在与物质相互作用时,光子只能作为一个整体吸收或产生。

在此理论下,光电效应的机制现在变得很清楚了。当光电材料内的电子吸收了一个光子,这光子的能量一部分用于电子脱离金属表面所需要的脱出功 $W$ 或功函数,余下的能量变为光电子的动能。这关系可表为

$$h\nu = \frac{1}{2}mv^2 + W \tag{10-2-1}$$

这个关系称为爱因斯坦光电方程。用它可以圆满地解释光电效应的实验事实。

注意到，爱因斯坦光电方程中，$W = h\nu_0$，$\nu_0$ 为红限波长所对应的频率，那么式(10-2-1)可改写成

$$\frac{1}{2}mv^2 = h(\nu - \nu_0) \tag{10-2-2}$$

式(10-2-2)表明，对于任何一种给定的材料，其动能对频率 $\nu$ 的关系是一条直线，斜率为普朗克常数，截距为逸出功 $W$，与频率轴的交点即为红限频率 $\nu_0$。

### 10.2.4　光的波粒二象性

**1. 德布罗意波(De Broglie wave)**

光作为波有两个重要的参数——波长 $\lambda$ 和频率 $\nu$，而作为粒子将用能 $E$ 和动量 $p$ 来描述，这两者之间的特征参数有如下关系：

$$\begin{aligned} E &= h\nu \\ \lambda &= \frac{h}{p} \end{aligned} \tag{10-2-3}$$

在光具有波粒二象性的启发下，1924 年，德布罗意在他的博士学位论文中指出，既然过去认为是波动的光具有粒子性，那么，作为实物粒子(指有静质量的粒子)也应该与光子一样，也具有波动性。因为自然界常常是对称的，对一个能量为 $E$ 和动量为 $p = mv$ 的实物粒子，对应的波的频率和波长为

$$\begin{aligned} \nu &= \frac{E}{h} \\ \lambda &= \frac{h}{p} = \frac{h}{mv} \end{aligned} \tag{10-2-4}$$

实物粒子的波称为德布罗意波或物质波。

由于 $h$ 很小，所以对宏观物体来说，其波动性很微弱，如一块 1g 质量的石子以 1m/s 的速度运动，波长仅 $7.6\times10^{-31}$m，这根本测不出。而对微观粒子，情况就不一样了，如一个由开始处于静止状态的电子越过 150V 的电位差后，其动能为 $eV = \frac{1}{2}mv^2$，其速度为 $v = \sqrt{2eV/m}$，其德布罗意波长为 $\lambda = \frac{h}{\sqrt{2eVm}} = 0.1$nm，这波长与 X 射线的波长相当了。

**2. 戴维孙-革末实验及其物理意义**

1928 年，德布罗意的物质波假说得到了证实，这是戴维孙(Dawisson)和革末(Gorrner)以及汤姆孙(Thomson)努力的结果。戴维孙、革末用镍单晶(面心立方结构)作为对电子衍射的三维光栅，当一束被加速到具有 54eV 能量的电子垂直于晶体 $M$ 的切割面入射时，在与法线成 50°的方向上出现一个很强的峰值，可以通过探测器 $G$ 进行测量，如图 10-2-2 所示。利用光栅的衍射公式

$$d\sin\theta = k\lambda \tag{10-2-5}$$

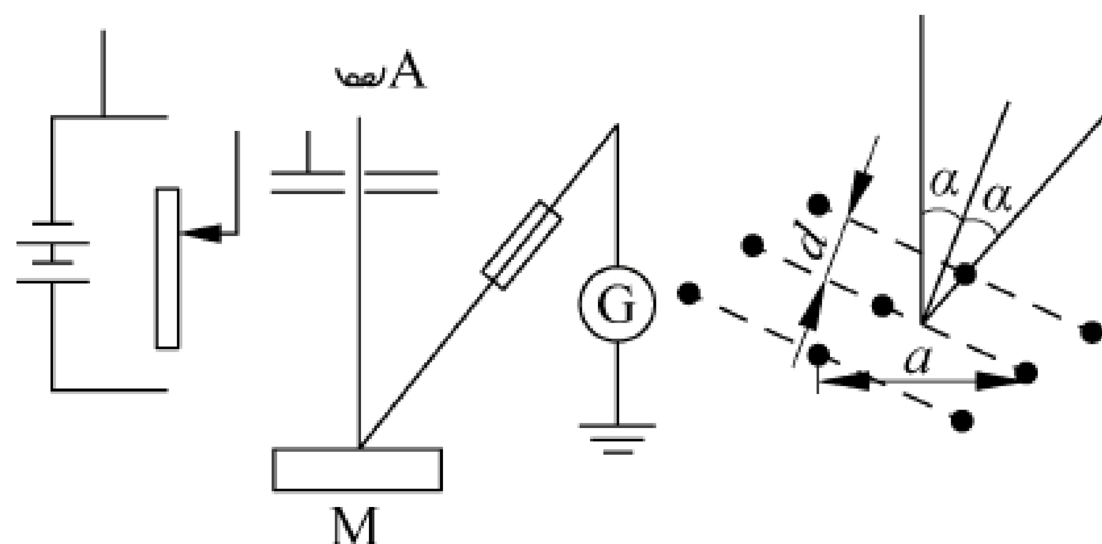


图 10-2-2　戴维孙-革末实验

进行实验和理论的对比研究。镍单晶间隔 $d=0.215\text{nm}$，$k=1$，得 $\lambda=0.165\text{nm}$，这与用式(10-2-5)计算结果 $\lambda=0.167\text{nm}$ 符合得很好。

汤姆孙则使用高速电子通过一张极薄的金属箔片，得到的衍射图样与用 X 射线所得到的图样完全类似，这样德布罗意预言的德布罗意波被实验所证实：实物粒子也具有波粒二象性。

德布罗意因德布罗意波的成果获得 1929 年诺贝尔物理学奖。

**3. 概率(概率)波**

光与电子都具有波粒二象性，如何理解波粒两者的关系呢。我们来看一个实验，分别用电子流和光照射一个狭缝。当电子流的密度很小，以致电子一个一个地通过狭缝，接收平面上开始时出现的是一些位置并不重合而且是无规则分布着的点，随着时间延长，点子数目增多最终形成衍射图样。模拟结果如图 10-2-3 所示。

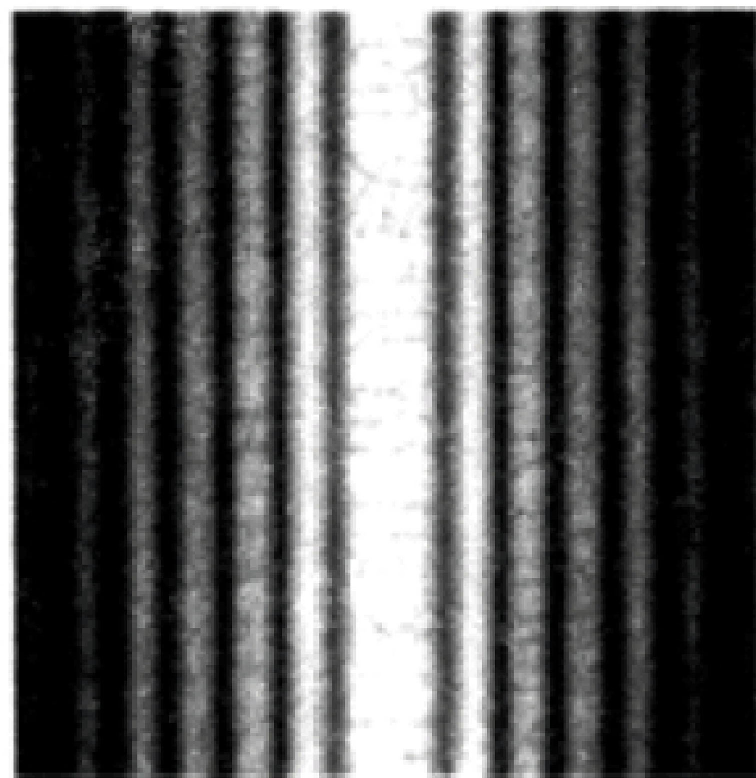


图 10-2-3 电子单缝衍射图样的形成

用光照射，当光强很弱时，照相底片上记录也是无规则的光点，随着曝光时间增加所得到的分布则显示衍射图样。如加大电子密度或光的强度，在短时间内也得到同样的衍射图样。这揭示了粒子性和波动性之间的关系，即单个光子和电子的行径是无规则的，而大量的光子和电子的分布与波动理论一致。此外，从统计的观点来看，大量的光子和电子同时通过与它们一个一个地通过之间的差别，只在于前一种是光子或电子下落到某一区域稠密些，后一种情况就是光子或电子落到这一区域频繁些，一个是对空间的统计平均，一个是对时间的统计平均。无论是哪种情况，从统计的角度可形成这样的概念：波在某时刻在空间某点的强度与该时刻在该点单位体积中出现光子或电子的概率成正比，强度大的地方出现光子或电子概率就大，强度小的地方出现光子或电子概率就小，这单位体积中的概率用 $\omega$ 表示，它可表示成一个复函数 $\psi$ 的绝对值的平方，即 $\omega=|\psi|^2$，我们称 $\psi$ 为波函数，光子或电子波动称为概率波。当光子和微观粒子处于不同的运动状态时，其波函数 $\psi$ 是不同的，由于概率波只给出波粒子在某处出现的概率，它给出了波粒子的粒子特征，它并不要求将波粒子分割，波粒子一旦出现在某处，总是整个粒子，即体现了它的粒子性。对光子，波函数满足量子化的麦克斯韦方程，对实物微粒，它满足薛定谔方程。这波函数就是我们对光子和微粒子具有波粒二象性特征的描述。

**4. 测不准关系式**

在经典力学中，为了确定宏观物体的运动状态，必须同时知道这个物体的位置(坐标)和动量。对于微观粒子，前面已知道它具有波粒二象性，那么还能用经典力学的位置(坐标)和动量来准确描述微观粒子的运动状态吗？

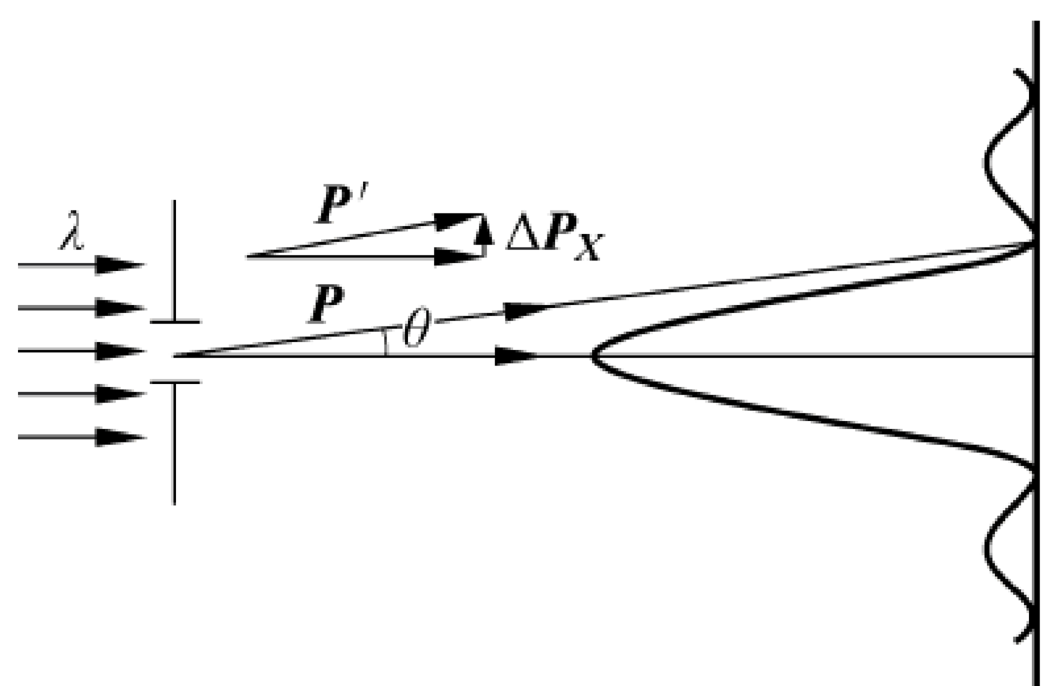



图 10-2-4 单缝的夫琅禾费衍射与测不准关系式

1927 年海森伯提出了测不准关系式，对这个问题作了否定的回答。为了说明这个问题，我们再次讨论狭缝的夫琅禾费衍射，如图 10-2-4 所示。

一束平行光入射到狭缝上，缝宽为 $d$，光子通过狭缝的位置不确定量为 $\Delta x=d$，通过狭缝将产生衍射，这可视作光子速度方向发生变化，引起动量的变化为

$p_x = p\sin\theta$，其中，$p$ 为光子的动量，它等于$\frac{h}{\lambda}$。所以，动量的不确定量为 $\Delta p_x = p_x$。考虑光子主要出现在主级大之内，有 $\sin\theta \sim \frac{\lambda}{d}$ 和 $\Delta p_x \sim \left(\frac{h}{\lambda}\right)\left(\frac{\lambda}{d}\right) = \frac{h}{d} = -\frac{h}{\Delta x}$，因此有 $\Delta p_x \Delta x \sim h$。如果考虑到光子主要出现在主级大之外，就有 $\Delta p_x \Delta x > h$。将此式推广到三维状态有

$$\begin{aligned} \Delta p_x \Delta x &> h \\ \Delta p_x \Delta y &> h \\ \Delta p_z \Delta z &> h \end{aligned} \tag{10-2-6}$$

这就是海森伯测不准关系式，它表明在同一方向上微观粒子的动量和位置不能同时准确确定，如果我们绝对准确地知道光子的动量，那么就不能确定它的位置，反之亦然，如果我们绝对准确地知道光子的位置，那么就不能确定它的动量。所以，这里的粒子性与经典概念的粒子是不同的，它的运动不能用轨道来描述。也就是说，不能知道这粒子究竟在哪一点，只能知道粒子在某一点出现的概率。

海森伯获得 1932 年诺贝尔物理学奖。

由上述讨论可知，黑体辐射、光电效应及康普顿效应等现象都说明光具有粒子性，只不过分别从光的发射和光与物质的相互作用不同方面揭示了光的粒子本性。至此，人们一方面通过光的干涉、衍射和偏振等光学现象证实了光的波动性；另一方面通过黑体辐射、光电效应和康普顿效应等又证实了光的粒子性。所以，为了将有关光的本性的两个完全不同的概念统一起来，人们进行了大量的探索工作。1924 年德布罗意创立了物质波学说，他大胆地设想每一物质的粒子都和一定的波相联系，这一假设在 1927 年被戴维孙和革末的电子束衍射实验所证实。事实上，不仅光具有波动性和微粒性，也就是所谓波粒二象性，而且一切习惯概念上的实物粒子同样具有这种二重性，也就是说，这是微观物质所共有的属性。1925 年玻恩提出的波粒的概率波概念，解释了波动性和粒子性之间的联系。

光和一切微观粒子都具有波粒二象性，推动人们去进一步探索光和物质的本质，包括实物和场的本质问题。这里的波、粒子已经不是经典理论中的那种概念，严格的表述将由量子电动力学给出。不过在一些宏观近似的条件下，如光的传播过程中，它可以过渡到用波动理论来描述；在另一些条件下，则显示出它的粒子性质，这种粒子性主要指在与物质相互作用时交换能量和动量的那种整体性。

## 10.3　激光原理与激光器

激光的英文全称是 light amplification by stimulated emission of radiation，缩写为 LASER，意思是受激辐射发出的光放大。汉语“激光”的名字是我国著名科学家钱学森命名的。

1958 年，肖洛（Arthur. L. Schawlow，1921—1999）和汤斯（Charles Hard Townes，1915—　）将这种微波激射器的原理应用到光频～$10^{14}$ Hz 波段，为激光器研发奠定了理论基础。1960 年，休斯(Hughes)在实验室制成并运转了第一台人造红宝石晶体激光器。几个月后，贝尔(Bell)电话实验室梅曼(Maiman，Theodore Harold，1927—2007)制成了第一台氦-氖气体激光器。从此，激光器的研制和各种激光技术的应用得到了突飞猛进的发展。激光器的种类已有固体、液体、气体、半导体和染料激光器。激光的波段从短至 0.24$\mu$m 的紫外，长到 774$\mu$m 的远红外，其中包括可见光、近红外等各个波段。输出功率从 $10^{-8}$～$10^{-12}$ N/s。出光方式也有连续和脉冲两种形式。激光技术的迅速发展，其形势可以同

20世纪50年代中半导体技术的发展相媲美,促进和开拓了一大批光学新的研究领域,如全息照相、光信息处理、非线性光学等。

### 10.3.1 激光产生的原理

激光是一种特殊的电磁波。激光的原理是100多年来科学家深入研究电现象、磁现象和光现象的结晶。激光原理的直接创始人,可以追溯到伟大的科学家爱因斯坦。1916年,爱因斯坦在《关于辐射的量子论》论文中提出原子中的电子可以“受激”而放出光子,这种受激辐射的过程就是产生激光的基本物理原理。

激光是原子受激辐射放大的简称。被限制在一个光学谐振腔内且处于激发态(低能级)的大量原子或分子,以一种有序的和高度相互依赖的方式同步地、一致地辐射,从而形成一种具有高度相干性的窄带光辐射——激光。

**1. 自发辐射与受激辐射**

激光的产生原理是利用了物质原子受激辐射后发生跃迁的特性。有两种原因可以产生这种电子的跃迁而导致原子发光。一种原因是原子内部自身的运动引起电子作自发激射产生跃迁,这种原因产生的高能态电子数较少,形成不了粒子数反转;另一种原因是来自外部或人为的原因。即通过外界的能量把大量的电子激发到高能级状态上去,形成粒子数反转,实现电子从高向低跃迁。

自发辐射是在没有任何外界作用下,激发态原子自发地从高能级 $E_2$ 跃迁到低能级 $E_1$,同时辐射出一光子,如图10-3-1所示,此时,发射的光子能量符合

$$h\nu = E_2 - E_1 \tag{10-3-1}$$

设发光物质单位体积中处于能级 $E_1$、$E_2$ 的原子数分别为 $N_1$、$N_2$,则单位时间内从 $E_2$ 向 $E_1$ 自发辐射的原子数为

$$\left(\frac{\mathrm{d}N_{21}}{\mathrm{d}t}\right) = A_{21}N_2 \tag{10-3-2}$$

$A_{21}$ 为自发辐射概率(自发跃迁率):表示一个原子在单位时间内从 $E_2$ 自发辐射到 $E_1$ 的概率。

受激吸收过程是原子吸收光子,从低能级 $E_1$ 跃迁到高能级 $E_2$,它对光放大起负作用,如图10-3-2所示,吸收的能量也是用式(10-3-1)描述。

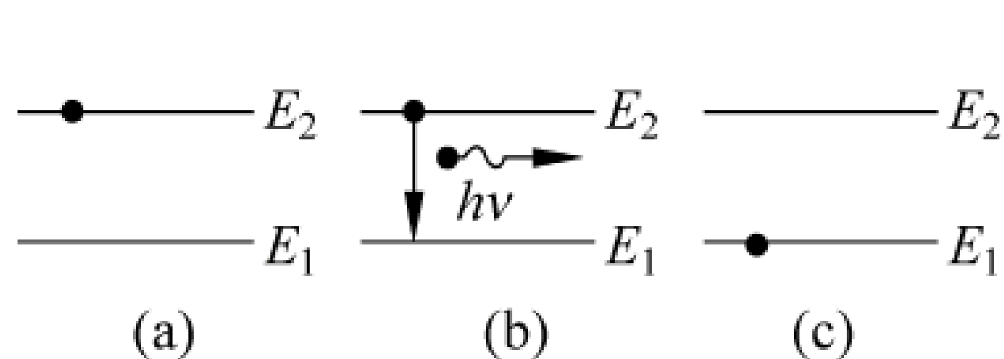


图10-3-1 光与原子的相互作用:自发辐射

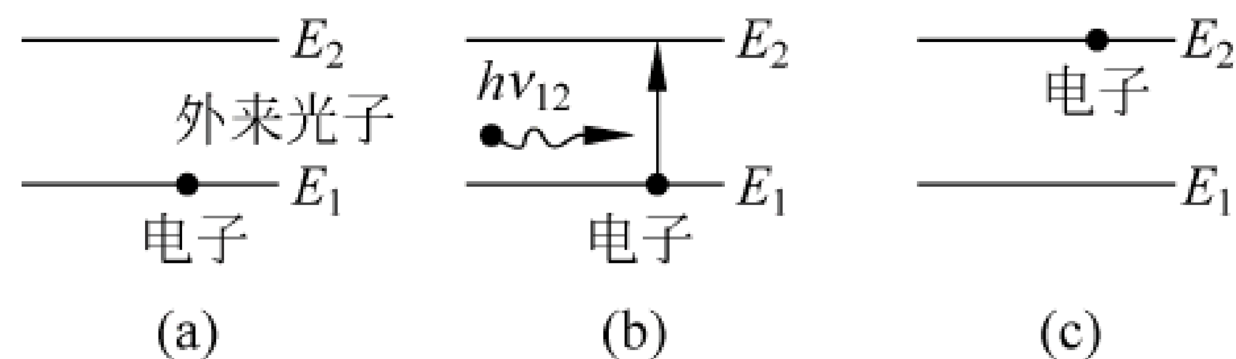


图10-3-2 光与原子的相互作用:吸收

受激辐射是原子在入射光的诱发下由高能级 $E_2$ 向低能级 $E_1$ 跃迁,同时放出一个与入射光相同的光子,即入射一个光子,出射两个光子,起着光放大的作用,这就是产生激光的物理机制,如图10-3-3所示。若体系中有大量原子处于某个激发态,其中有一个原子自发发射光子,这个光子会诱发处于该能级的其他原子产生受激辐

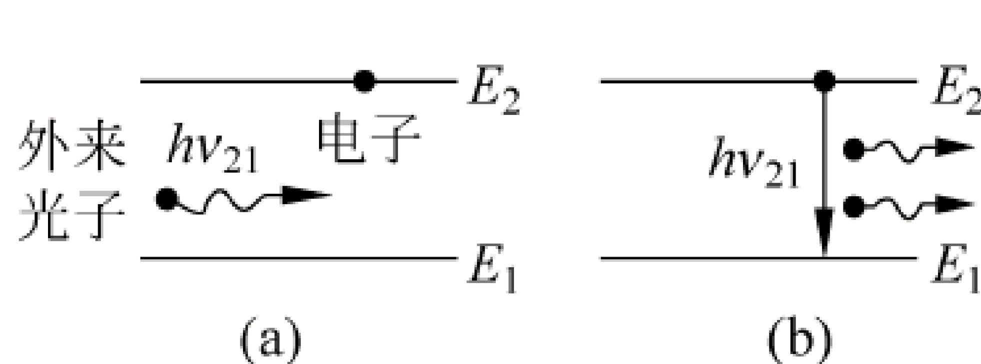


图10-3-3 光与原子的相互作用:受激辐射

射，发射光子。这些光子又会使另一些原子受激发射光子从而引起连锁反应，产生强烈的激光。受激辐射是引起光放大、产生激光的主要原因。

**2. 产生激光的两个必要条件**

产生激光的首要必要条件是必须实现粒子数反转。受激吸收原子数与处于 $E_1$ 的原子数 $n_1$ 成正比，受激辐射原子数与 $E_2$ 的原子数 $N_2$ 成正比。当 $N_2<N_1$ 时发生受激辐射远少于发生受激吸收，是不可能实现光放大的。要实现光放大，必须采取特殊措施，打破原子数在热平衡下的玻耳兹曼分布，使 $N_2>N_1$。我们称体系的这种状态为粒子数反转。可见，并不是所有的物质都能实现粒子数反转。能够实现粒子数反转的介质称为激活介质。要造成粒子数反转分布，首先要求介质有适当的能级结构，其次还要有必要的能量输入系统。供给低能态的原子以能量，使物质中有尽可能多的粒子吸收能量后，从低能级不断跃迁到高能级上去，这一能量供应过程，叫作"激励""激发"，或者叫作"抽运""泵浦"等。

产生激光的第二必要条件是受激辐射要远大于自发辐射。在一般的可见光源里，自发跃迁概率远大于受激跃迁概率，所以为使受激辐射概率远大于自发辐射概率，在激光器中可用两个高反射率的镜面组成谐振腔，使辐射在腔内反复振荡，大大提高辐射场的能量密度，从而使受激辐射概率远大于自发性辐射概率，所以，谐振腔是激光器必不可少的器件。

**3. 激光的特性**

激光是20世纪以来，继原子能、计算机和半导体之后人类的又一重大发明。激光是在有理论准备和生产实践迫切需要的背景下应运而生的，它一问世，就获得了异乎寻常的飞快发展，激光的发展不仅使古老的光学科学和光学技术获得了新生，而且导致许多新兴产业的出现。那么，激光为什么有如此大的魅力呢？让我们来看看它与普通光到底有什么不同。

第一，激光的单色性好。比如，单模氦-氖激光器发出波长为632.8nm的激光，其谱线宽度 $\Delta<10^{-6}$ nm。激光是一种颜色最单纯的光。太阳光和电灯光看起来似乎是白色的，但当让它通过一块三棱镜的时候，就可以看到红、橙、黄、绿、青、蓝、紫七种颜色的光，其实，还含有我们看不见的红外光和紫外光。

第二，激光的方向性好。在发射方向的空间内光能量高度集中，所以激光的亮度比普通光的亮度高千万倍，甚至亿万倍。而且，由于激光可以控制，使光能量不仅在空间上高度集中，同时在时间上也高度集中，因而可以在一瞬间产生出巨大的光热，成为无坚不摧的强大光束。平时，我们见到的灯光，都是向四面八方发光。而激光却不同，它是大量原子由于受激辐射所产生的发光行为。激光在传播中始终像一条笔直的细线，发散的角度极小。

第三，激光亮度最高。激光的出现，是光源亮度上的一次飞跃。一台普通的激光器的输出亮度，比太阳表面的亮度大10亿倍。

第四，激光的相干性好。激光的相干长度达数十公里，而最好的普通光源氪灯的相干长度也只有数十厘米。

第五，激光还可以具有很大的能量，具有广泛的应用领域。在工业生产中，利用激光高亮度特点已成功地进行了激光打孔、切割和焊接；在医学上，利用激光的高能量可剥离视网膜凝结和进行外科手术；在农业技术上，激光可以提高育种效率；在测绘方面，可以进行地球到月球之间距离的测量和卫星大地测量；在军事领域，激光可以制成摧毁敌机和导弹的激光武器；等等。

### 10.3.2　激光器

一般激光器应由工作物质、激励光源（光泵）和光学谐振腔三部分组成，如图10-3-4

所示。

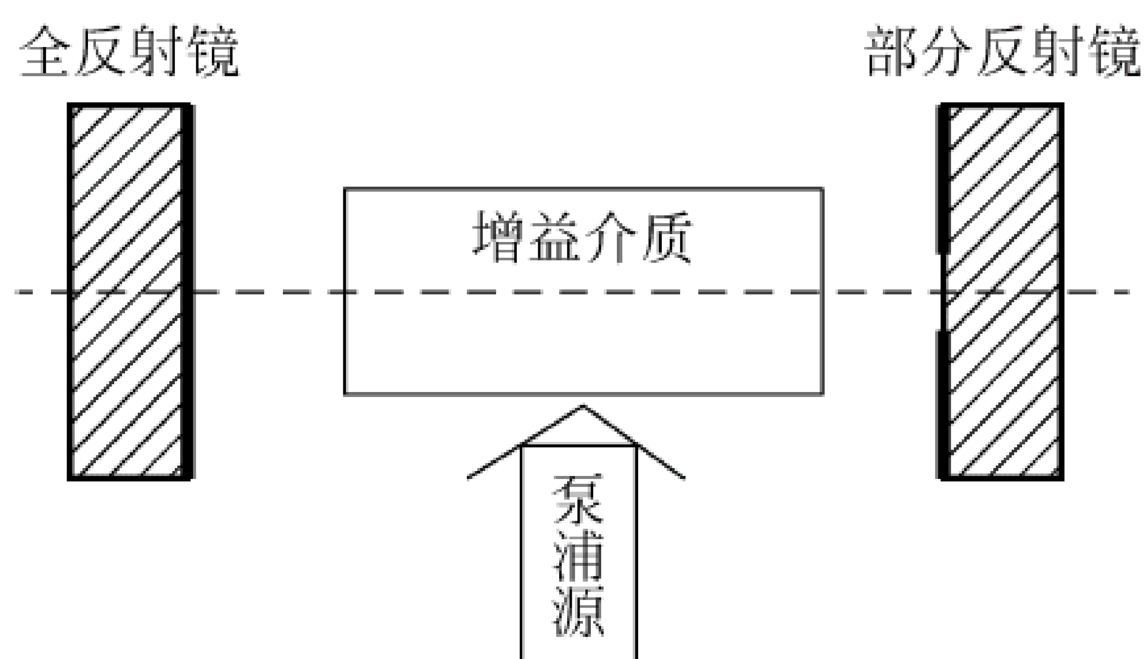


图 10-3-4 激光器的构成

目前激光器的种类很多,对不同的激光器有不同的分类方法。按工作物质的不同可以分为固体激光器、气体激光器、液体激光器和半导体激光器;根据激光输出方式的不同又可分为连续激光器和脉冲激光器,其中脉冲激光的峰值功率可以非常大;还可以按发光的频率和发光功率大小分类。气体激光器可以是原子气体激光器,也可以是分子气体或离子气体激光器。这里只选择几个有代表性的激光器作一简单介绍。

(1) 固体激光器

一般讲,固体激光器具有器件小、坚固、使用方便和输出功率大的特点。这种激光器的工作介质是在作为基质材料的晶体或玻璃中均匀掺入少量激活离子,在固体中能产生受激发射作用的金属离子主要有三类:第一类是过渡金属离子(如 $Cr^{3+}$);第二类是大多数镧系金属离子(如 $Nd^{3+}$、$S^{2+}$、$Dy^{2+}$ 等);第三类是锕系金属离子(如 $U^{3+}$)。这些掺杂到固体基质中的金属离子的主要特点是:具有比较宽的有效吸收光谱带,比较高的荧光效率,比较长的荧光寿命和比较窄的荧光谱线,因而易于产生粒子数反转和受激发射。

用作晶体类基质的人工晶体主要有:刚玉($Al_2O_3$)、钇铝石榴石($Y_3Al_5,O_{12}$)、钨酸钙($CaWO_4$)、氟化钙($CaF_2$)以及铝酸钇($YAlO_3$)、铍酸镧($La_2Be_2O_5$)等。用作玻璃类基质的主要是优质硅酸盐光学玻璃,例如,常用的钡冕玻璃和钙冕玻璃。与晶体基质相比,玻璃基质的主要特点是制备方便和易于获得大尺寸优质材料。对于晶体和玻璃基质的主要要求是:易于掺入起激活作用的发光金属离子;具有良好的光谱特性、光学透射率特性和高度的光学(折射率)均匀性;具有适于长期激光运转的物理和化学特性(如热学特性、抗劣化特性、化学稳定性等)。晶体激光器以红宝石($Al_2O_3$:$Cr^{3+}$)和掺钕钇铝石榴石(简写为 YAG:$Nd^{3+}$)为典型代表。玻璃激光器则是以钕玻璃激光器为典型代表。

(2) 气体激光器

气体激光器具有结构简单、造价低,操作方便,工作介质均匀,光束质量好以及能长时间较稳定地连续工作的特点。这也是目前品种最多、应用最广泛的一类激光器,市场占有率达60%左右。

以气体为工作物质,常见的有氦氖、氦镉离子、氩离子和二氧化碳气体,这种激光器通常用气体放电激励,特殊情况才采用光激励等其他激励方式,气体激光器光束质量比较好,出光方式一般为连续式。



图 10-3-5 氦氖激光器的结构

氦氖激光器:常见结构如图 10-3-5 所示,主要输出波长为 632.8nm、1152.3nm 和 3391.3nm,输出功率 1~100mW。

氦镉离子激光器:主要输出波长为 325nm、416nm,输出功率几十毫瓦。

氩离子激光器:主要输出波长为 488nm、514.5nm,输出功率为瓦的数量级。

二氧化碳激光器:输出波长为 10~600nm 红外光,输出功率数百瓦甚至数十万瓦的数量级。

(3) 液体激光器

常用的是染料激光器,采用有机染料作为工作介质。大多数情况是把有机染料溶于溶

剂中(乙醇、丙酮、水等)中使用,也有以蒸气状态工作的。利用不同染料可获得不同波长激光(在可见光范围)。染料激光器一般使用激光作泵浦源,例如,常用的有氩离子激光器等。

液体激光器工作原理比较复杂。染料激光器的波长调节范围可从紫外的 300nm 到近红外的 1.2μm。在各种染料系列中,经常使用的是若丹明,这是一种红色染料,输出激光能量最高,转换效率也较高,调谐范围在 500～700nm。因为染料激光器输出波长连续可调,且覆盖面宽,使它也得到广泛应用。

(4) 半导体激光器

半导体激光器是以半导体材料作为工作介质的。目前较成熟的是砷化镓激光器,发射 840nm 的激光。另有掺铝的砷化镓、硫化铬硫化锌等激光器。激励方式有光泵浦、电激励等。这种激光器体积小、质量轻、寿命长、结构简单而坚固,特别适于在飞机、车辆、宇宙飞船上用;另外在作扫描光源时,不需要电光或声光调制器,因为它有很好的时间响应特性,而且可以通过调制注入电流来调制输出激光强度,室温下的响应时间在毫微秒量级,内调制频率可达兆赫。它们的光功率在 5～35mW。近年来,由于光纤通信和光盘技术的发展大大推动了半导体激光器的发展。

(5) 自由电子激光器

X 射线的波长范围是 0.01～70nm,然而在这个区域实现激光发射的技术要求很高,因为增益随波长变短而迅速下降,谐振腔反射镜效率极低而所需泵功率极高。获得相干 X 射线的方法有:谐振混频法、γ 射线激光、高温等离子体激光和自由电子激光。无论哪一种方法,目前都还处于实验室研究阶段。

自由电子激光器不存在使用寿命问题,也可避免一般激光器的某些工艺上的麻烦,如激光工作物质稀缺或有毒。当然,整个激光设备庞大昂贵,还无法广泛应用。

## 10.4 非线性光学与现代光学技术

非线性光学是自 20 世纪 60 年代激光出现之后,迅速发展起来的光学中的一门新的科学分支。非线性光学的出现,很快地引起了许多科学工作者极大兴趣,进行了大量的实验和理论研究工作。特别 20 世纪 70 年代以来,借助于可调频激光器和脉冲激光器,观察到了许多新现象,进一步促进了非线性光学的发展。

### 10.4.1 非线性光学

在通常的光学效应中,当光波穿过介质时,光波的电场强度 $\boldsymbol{E}$ 及其在介质中引起的极化强度 $\boldsymbol{P}$ 是成线性关系,即 $\boldsymbol{P}=X\boldsymbol{E}$。此时表征物质性质的许多光学参数,如吸收系数,折射率,散射截面等都是与场强无关的常量。这是线性光学的重要特点。所谓非线性光学效应,就是物质在很强的激光作用下,介质的电极化强度 $\boldsymbol{P}$ 和入射光波的电场强度 $\boldsymbol{E}$ 之间的关系,不再遵循线性关系。也就是说,在介质的极化强度 $\boldsymbol{P}$ 和光波性电场强度 $\boldsymbol{E}$ 的关系中,除了含有原线性项之外,还有光波电场的平方项、三次方项以及更高次项,即

$$\boldsymbol{P}=x\boldsymbol{E}+\alpha_1 x\boldsymbol{E}+\alpha_2 x\boldsymbol{E}+\cdots \tag{10-4-1}$$

正是这些非线性项产生了一系列的非线性光学效应。在已知非线性光学效应中大体可分两类:一类是二阶的非线性光学效应;另一类是三阶非线性光学效应。

典型的二阶非线性光学效应有:

(1) 光学的二次谐波(或称光学倍频)频率为 ω 的激光作用于非线性介质中,在位相匹

配的条件下,就会产生频率为 $2\omega$ 的激光。

(2) 光学的混频(或称和频与差频)两个频率分别为 $\omega_1$ 和 $\omega_2$ 的激光,同时作用在非线性的介质,在一特定的条件下,就会产生频率为 $\omega_1+\omega_2$ 和 $\omega_1-\omega_2$ 的激光。

(3) 光的参量放大和参量振荡,一束频率为 $\omega_n$ 的所谓泵浦光和另一频率为 $\omega_m$ 的信号光,同时作用于非线性晶体上,当晶体相对于光传播方向有适当的取向时,泵浦光就可转变为放大了的光信号,同时产生 $\omega_i=\omega_n-\omega_m$ 的闲频光,如果对信号再加上一个光学谐振腔,则对信号放大变成了信号频率的自频振荡,此时,虽然没有信号输入,但会有自动产生信号光,同时,产生闲频光。当晶体的取向或温度发生改变时,信号光的闲频率都随之改变。

非线性光学效应只能在不具有对称中心的各向异性晶体中才能发生,要使非线性光学现象得到迅速的应用,有许多问题需要解决,但是其中最重要的是要有良好的光学材料,对非线性光学材料的要求是:有高的非线性极化系数,可使用的频带宽且透明性好,能以某种方式实现相位匹配;能承受光强照射,不易产生光学损伤等。目前发展的重点是非线性光学晶体,常见的晶体有三类:磷酸或砷酸二氢碱金属盐简称类 KDP,如 $KH_2OP_4$;碘酸碱金属盐,如 $LiIO_3$;妮酸盐,如 $LiNbO_3$ 等。最近发展比较快的是妮酸钡钠(BNN)。

利用上述非线性光学效应制成光的倍频、光的变频器、光的参量放大和参量振荡器,在激光的领域中有广泛应用,并不断取得新的进展。有力地促进了激光技术的发展,反之也为非线性光学的发展创造了可靠的条件,因此,无论从理论上还是从实用上研究非线性光学都有重要意义。非线性光学有极其光明的发展前景。

**1. 激光倍频技术**

当入射到介质中的光波 $E=E_0\cos\omega t$ 很强时,则晶体中产生的电极化强度为

$$P=\alpha E+\beta E^2=\alpha E_0\cos\omega t+\beta E_0^2\cos^2\omega t \tag{10-4-2}$$

根据三角公式,上式可写成

$$\begin{aligned}P&=\alpha E_0\cos\omega t+\frac{1}{2}\beta E_0^2(1+\cos 2\omega t)\\&=\frac{1}{2}\beta E_0^2+\alpha E_0\cos\omega t+\frac{1}{2}\beta E_0^2\cos 2\omega t\end{aligned} \tag{10-4-3}$$

从式(10-4-3)可以看出,电极化强度除了有直流成分外,还有频率为 $\omega$ 的基频成分以及频率为 $2\omega$ 的倍频成分,由于这些电极化强度的存在,就相应地产生了基频极化波 $P(\omega)$ 和倍频极化波 $P(2\omega)$,而 $P(\omega)$ 和 $P(2\omega)$ 又产生相应的基频次波辐射 $E'(\omega)$ 和倍频次波辐射 $E'(2\omega)$。这就是倍频光产生的机理,如图 10-4-1 所示。

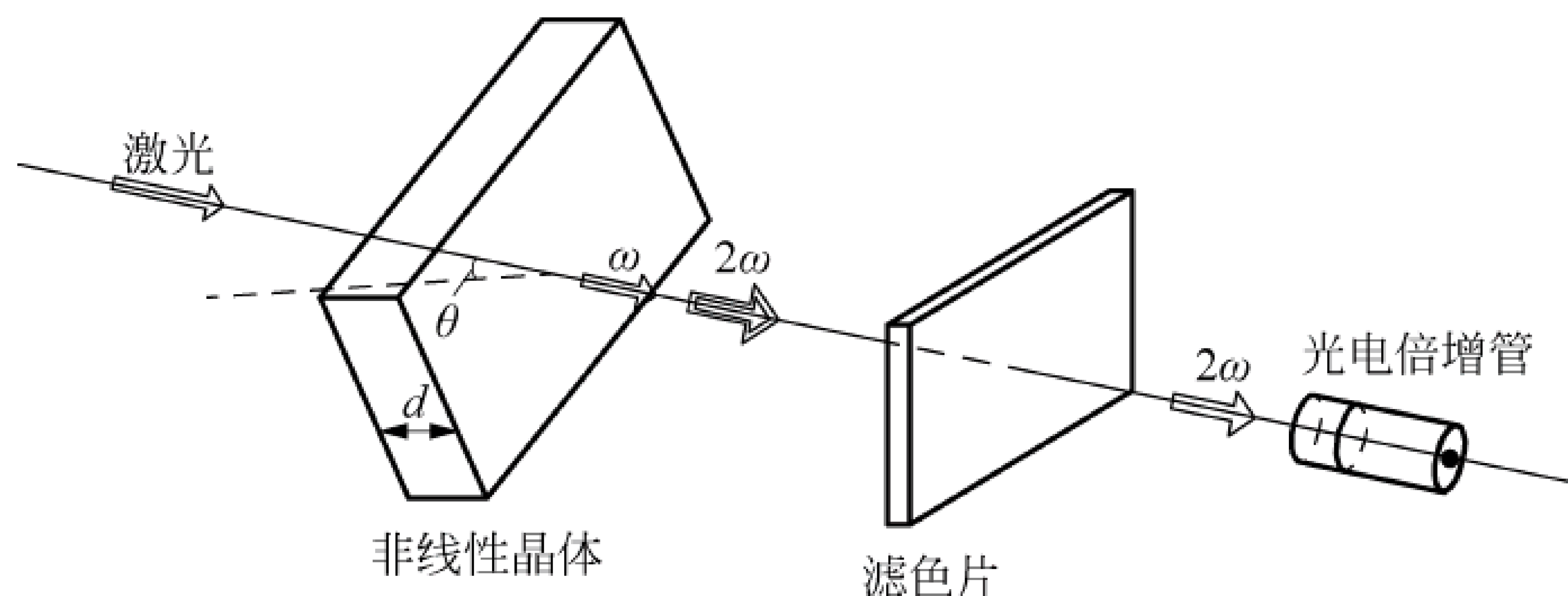



图 10-4-1 倍频光的产生

**2. 受激喇曼散射**

强光与激活介质相互作用的非线性光学现象中,有一类是受激辐射。当光束通过光学

性质不均匀的物质时，从它的侧面也可以看到光，这种现象叫作光的散射。在通常的情况下，散射光的频率与入射光的频率相同，这种散射现象称为瑞利散射。在各向异性的物质中，瑞利散射的频率也稍有变化，在一些特殊情况下，在散射光中，除了含有与入射光频率 $\nu_0$ 相同的光外，还有频率为 $\nu_0 \pm \Delta\nu$ 的光存在。这里 $\Delta\nu$ 与物质的性质有关，而与入射光的频率 $\nu_0$ 无关，这种散射现象称为喇曼（Sir Chandrasekhara Venkata Raman，1888—1970）散射。

散射过程的原理是这样的：在入射光的作用下，物质分子吸收一个入射光子后便跃迁到一个特殊的能级上，当这个分子从该能级跃迁，回到原来的能级时，能发射出一个与入射光频率相同的散射光子，这就是瑞利散射。当该能级上的分子跃迁到比原来能级低或高的能级时，将发射出与入射光频率不同的散射光子，向低频方向移动的散射光谱线 $\nu_0 - \Delta\nu$ 叫作斯托克斯（G. G. Stokes，1819—1903）线；向高频方向移动的散射光谱线 $\nu_0 + \Delta\nu$ 叫作反斯托克斯线，这里的 $\Delta\nu$ 只与物质的结构有关。当以强激光入射时，可使某些介质的散射过程具有受激发射的性质，即散射光突然变强，超过原来的几百倍到上千倍，光谱线变窄，并且显示出与激光同样的方向性，此种现象为受激散射。

**3. 激光自聚焦**

在激光的横模光斑图中，光强分布是不均匀的，即使基模光斑中不存在光强为零的场点，能量也集中在中心，且按高斯函数规律由中心向外平滑地减小，这种在截面内光强分布不均匀的光束，在通过非线性介质时，会引起介质折射率感应变化不均匀，从而导致激光自聚焦，所以，激光自聚焦是一种感应透镜效应。

设有一单模激光束，它具有高斯函数型的横向分布，在非线性介质中传播时的折射率 $n$ 由两部分组成：

$$n = n_0 + \Delta n(|E^2|) \tag{10-4-4}$$

式(10-4-4)中前一项 $n_0$ 为普通的折射率；后一项与 $|E^2|$ 成正比，是非线性折射率，$\Delta n$ 为光场感应引起的折射率变化。如果 $\Delta n$ 是正的，则对高斯横向分布的激光束来说，中心部分折射率比边缘部分折射率大，于是激光束好像通过一个正透镜一样，产生会聚作用。

强激光的自聚焦会导致光学元件破损，防止的办法是尽量设法使横向光强分布均匀，通常采用发射光或准平行光入射，以减小介质折射率的不均匀程度。

## 10.4.2　现代光学技术

**1. 激光测距技术**

随着近代光学的飞速发展，光学技术已渗透到各个学科。尤其是激光技术，现在激光已广泛应用到医疗、工业、环保等各个领域并发挥着不可替代的作用。我们从激光的原理等方面出发去探讨其在测量中的应用。

激光为定向辐射，其平面发散可从几毫弧度至十几毫弧度，具有很高的准直性。如再经发射望远镜发射出去，其平面发散角一般可控制在十分之几毫弧度以内。激光的这一特点使发射到大气中的激光光束，在十公里远处的光斑直径仅扩展到几米至几十米。而普通光源则向整个空间辐射，如要获得平行光束，则需要将光源置于反射镜或透射镜的焦点上。然而，由于光源具有一定的尺度，只能将光源的部分能量转换为准直光束。对光束的平行度要求愈高，被利用的光源能量也就愈少。因此，激光具有较高的准直性，可以获得极强的发光亮度和较远的探测距离。所以，激光是用于测距的最佳光源，而方向性也就成为激光测距必不可少的条件。

(1) 相位式激光测距仪

激光测距是光波测距中的一种测距方式,如果光以速度 $c$ 在空气中传播,在 $A$、$B$ 两点间往返一次所需时间为 $t$,则 $A$、$B$ 两点间距离 $D$ 可用下式表示:

$$D=\frac{ct}{2} \tag{10-4-5}$$

式(10-4-5)中:$D$ 为测站点 $A$、$B$ 两点间距离;$c$ 为光在大气中传播的速度;$t$ 为光往返 $A$、$B$ 一次所需的时间。

由式(10-4-5)可知,要测量 $A$、$B$ 距离实际上是要测量光传播的时间 $t$,根据测量时间方法的不同,激光测距仪通常可分为脉冲式和相位式两种测量形式。

相位式激光测距仪是用无线电波段的频率,对激光束进行幅度调制并测定调制光往返测线一次所产生的相位延迟,再根据调制光的波长,换算此相位延迟所代表的距离。即用间接方法测定出光经往返测线所需的时间,如图 10-4-2 所示。相位式激光测距仪一般应用在精密测距中。由于其精度高,一般为毫米级,为了有效地反射信号,并使测定的目标限制在与仪器精度相称的某一特定点上,对这种测距仪都配置了被称为合作目标的反射镜。

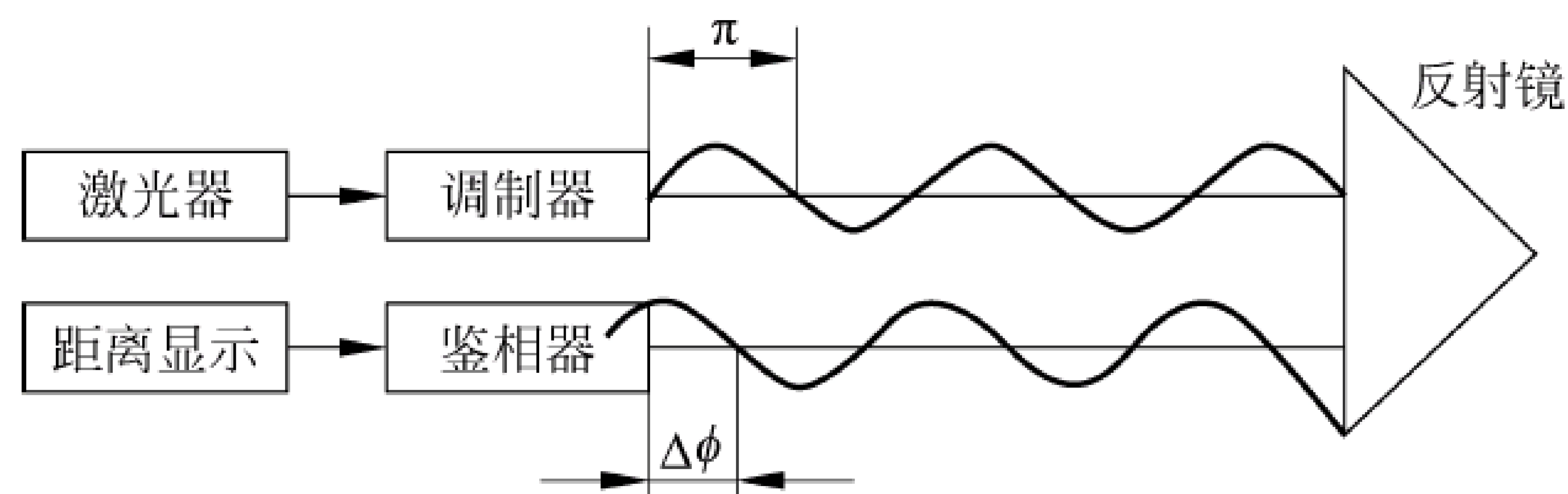



图 10-4-2 相位式激光测距原理

若调制光角频率为 $\omega$,在待测量距离 $D$ 上往返一次产生的相位延迟为 $\varphi$,则对应时间 $t$ 可表示为

$$t=\frac{\varphi}{\omega} \tag{10-4-6}$$

将此关系代入式(10-4-5),则距离 $D$ 可表示为

$$D=\frac{1}{2}ct=\frac{1}{2}\cdot\frac{c\varphi}{\omega}=\frac{c}{4\pi f}(N\pi+\Delta\varphi)=\frac{c}{4f}(N+\Delta N) \tag{10-4-7}$$

式中,$\varphi$ 为信号往返测线一次产生的总的相位延迟;$\omega$ 为调制信号的角频率,$\omega=2\pi f$;$N$ 为测线所包含调制半波长个数;$\Delta\varphi$ 为信号往返测线一次产生相位延迟不足 $\pi$ 部分;$\Delta N$ 为测线所包含调制波不足半波长的小数部分。

$$\Delta N=\frac{\varphi}{\omega} \tag{10-4-8}$$

在给定调制和标准大气条件下,频率$\frac{c}{4\pi f}$是一个常数,此时距离的测量变成了测线所包含半波长个数的测量和不足半波长的小数部分的测量,即测 $N$ 或 $\varphi$。由于近代精密机械加工技术和无线电测相技术的发展,已使 $\varphi$ 的测量达到很高的精度。为了测得不足 $\pi$ 的相角 $\varphi$,可以通过不同的方法来进行测量,通常应用最多的是延迟测相和数字测相,目前短程激光测距仪均采用数字测相原理来求得 $\varphi$。

(2) 对卫星进行的激光测距技术

从地面测量站向卫星发射一束激光,经过卫星上的激光反射器反射后,由测量站的接收

设备接收，测出激光往返的时间间隔，从而推算出卫星的距离，如图 10-4-3 所示。

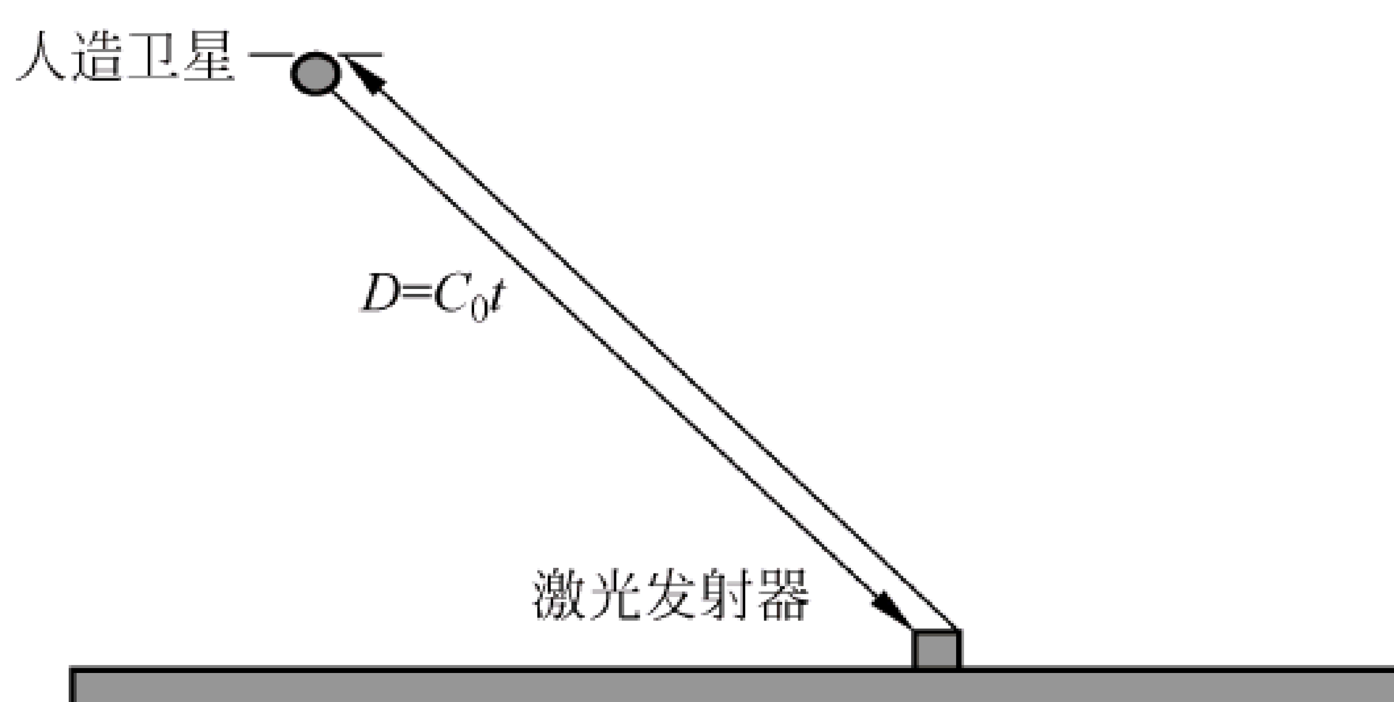



图 10-4-3　卫星激光测距技术

激光测距可分为脉冲测距和连续波相位测距两种。前者测量光脉冲在待测距离上往返传播的时间间隔；后者测量光束上调制信号在待测距离上往返传播时所发生的相位变化，间接测量时间间隔，得到卫星的距离。进行激光测距的卫星上必须安装若干个激光角反射器，它是一种光学四面体棱镜，用以增大反射能量，并使反射光沿着与入射光平行的方向反射回地面的激光测距仪，从而可增大测量距离。卫星是高速运动的目标，为了使窄激光束对准卫星目标，激光测距仪必须安装在精密的方向跟踪架上或电影经纬仪和卫星跟踪摄影机上，接收回波就可以达到既测距又测角，实现单台定位。

此项技术使人类身处在地球表面，却可时时把握人造卫星的位置情况，以便作不断的调整。相应地，人们依靠相同的原理，测出宇宙间各个星球离地球的距离，为天文学的研究工作提供了极其重要的便利。

（3）激光高度计

激光高度计的主要工作方式是利用计算发射和接收到激光的时间差来进行距离的测量。它以其高精确度、高分辨率受到工程师们的青睐，并被广泛地应用于遥感、航空航天等领域。

激光高度计可被安装于飞机、卫星等测试仪器平台上，它由激光发射模块、激光接收模块和数据处理模块三部分组成。激光发射模块发射出的激光首先打在地面、洋面上，激光接收模块收到反射回来的光信号，并把它转化为电信号。数据处理模块会精确地测量出从激光发射到激光高度计接收到激光的时间，而这段时间就是激光在大气中的传输时间。在这段时间内，激光行走的路程是高度计与测量目标间距离的两倍。

根据光在空气中的传播速度，可以计算出这个距离的大小。再根据激光高度计的空中高度，就可以最终得到探测目标的海拔高度。由于激光束具有较小的发散角，因此，激光束打在目标上会形成较小的光斑直径，这使得激光高度计有很好的水平分辨率。如果我们高密度地获得探测点，就能够得到较精密的探测区域的地形图。

美国科学家在第一代和第二代火星探测器上都采用了这种激光高度计来获得火星表面的地形分布图。科学家们采用不同的颜色来标定不同的高度，火星表面的地貌就简洁形象地展现在我们面前。这无疑为人类探测宇宙提供了强有力的武器，也为火星探测器选定最佳着陆地点提供了有价值的资料。

在激光高度计对地球表面的探测领域中，美国科学家们也曾多次应用激光高度计来勘探地表的植被覆盖情况、海洋面的高度等地表的物理特性以及云层和高空中悬浮物质的光学厚度。通过安装在飞机或卫星上激光高度计的帮助，人类探索的触角伸向了地球更深处。

现在我国的航天科学家们在致力于第一阶段“探月一号”卫星的技术与研究。在深入分析、研究国际月球探测的发展和已取得的成果的基础上，有关专家结合我国月球探测的发展计划和技术基础，经过充分的论证，确定了第一阶段“探月一号”卫星的科学和应用目标。这其中，激光高度计将实现获取卫星下方月球地形高度数据的任务，服务于月球地形高度数据

的任务,服务于月球表面三维影像获取的科学和应用目标。通过星上激光高度计测量卫星到星下点同月球表面的距离,为光学成像探测系统的立体成图提供了修正参数,并通过地面应用系统将距离数据与卫星轨道参数,地月坐标系进行综合数据处理,获取卫星下点同月表地形高度数据。

"探月一号"卫星的工作阶段分为发射轨道阶段、转移轨道阶段和环月轨道阶段,激光高度计在进入环月轨道之前不工作,在进入环月阶段之后,不论月球表面白天和黑夜,也不论卫星处于正飞或侧飞状态,激光高度计长期开机工作,因此,激光高度计的设计不仅要注重参数指标的实现,也要努力提高其工作稳定和持久性。

**2. 运用激光测量大气指数**

激光的这一系列特性对于激光探测大气污染也是至关重要的。在讲述其工作原理之前先简单了解一下现在大气污染的概况也是相当必要的。

人类活动造成的大气污染主要来自燃烧、工业生产过程及交通运输过程等带来的污染物。主要有燃烧时产生的烟尘、$SO_2$、CO、氮氧化合物和碳氢化合物等。其次是生产过程中排放的 Pb、Hg、Cd 等重金属粉末,以及 $Cl_2$、$H_2S$、$HNO_3$、$H_2SO_4$ 和氟化物等有毒气体。这些污染物虽然数量有限,但毒性较强,危害较大,往往引起严重的局部污染。最后汽车排出的废气,由于含有大量的 CO、氮氧化合物、碳氢化合物和铅化物等有害物质,并且在地面附近排放,因此对人体的危害较大。当这些污染物在太阳紫外光的照射下,还会产生光化学反应,形成光化学烟雾,造成危害较大的二次污染。

那么,激光是怎样对这些污染物进行探测的呢?下面让我们来一一探讨。

(1) 激光探测大气气溶胶

广义上说,大气气溶胶就是指悬浮在大气中的各种固态和液态微粒,如烟灰、尘埃、海盐、云雾和降水等微粒的总称。不过习惯上大气气溶胶不包含云雾、降水粒子。在一定条件下,利用单波长激光雷达所探测气溶胶消光系数的时空分布,可获得气溶胶浓度或质量浓度的时空分布。通常,在能见度不太恶劣的天气条件下,气溶胶对激光的散射满足一次散射和独立散射。气溶胶数浓度 $N\alpha(R)$ 和质量浓度 $M\alpha(R)$,均正比于气溶胶消光系数 $\sigma(R)$,并可分别表示为

$$N\alpha(R)=\xi_1(R)\sigma(R) \tag{10-4-9}$$

$$M\alpha(R)=\xi_2(R)\sigma(R) \tag{10-4-10}$$

式中 $\xi_1(R)$ 和 $\xi_2(R)$ 为比例系数,与激光波长、气溶胶归一化谱分布函数和光学折射率有关。因此,若激光探测时气溶胶物理特性的空间分布较为均匀,则探测空间的比例系数近似常数。在此条件下,根据激光雷达所探测地面平均大气消光系数,以及用取样法同时观测的地面气溶胶数浓度或质量浓度,便可获得比例系数的地面值 $\xi_1(0)$ 和 $\xi_2(0)$,并可将其推广至整个探测空间。于是,根据激光雷达所探测到的大气气溶胶消光系数的空间分布,由式(10-4-9)或式(10-4-10)可进一步获得气溶胶数浓度或质量浓度的空间分布。

(2) 激光探测烟云

激光可探测到几公里远,甚至肉眼已看不清楚的烟云。这对直观了解表征烟云形态的几何形状、上升轨迹、扩散范围以及它们随时间的变化等,提供了一种有效的探测手段。根据激光雷达探测的烟云回波波形,可获得激光传输路径上的烟云宽度。由激光雷达的仰角、方位角和烟云回波斜距,经过简单的几何运算,可获得激光传输路径上所探测烟云的空间位置。

其探测烟云上升轨迹和扩散范围的具体方法可简述如下。若近地层大气无风切变，以及中性或不稳定大气条件下，烟云横风向扩散一般呈轴向对称分布。这时，利用激光雷达只需探测离烟源下风不同距离烟云轴线处的烟云回形波波形，便可由此确定烟云上升轨迹和烟云横风向扩散范围，而不必探测烟云横风向垂直剖面。

图 10-4-4 表明，从烟囱口排放到大气中的烟云，由于热力和动力的抬升作用，迅速上升至 200～300m 的高度。然后，烟云一方面随气流向下风方扩散，另一方面继续上升，直至离烟源下风方 1km 以外，烟云上升才渐趋缓慢。至 4km 处，烟云高度趋于稳定，此时烟云轴线高度达 450m 左右，烟云横风向宽度则接近 300m。

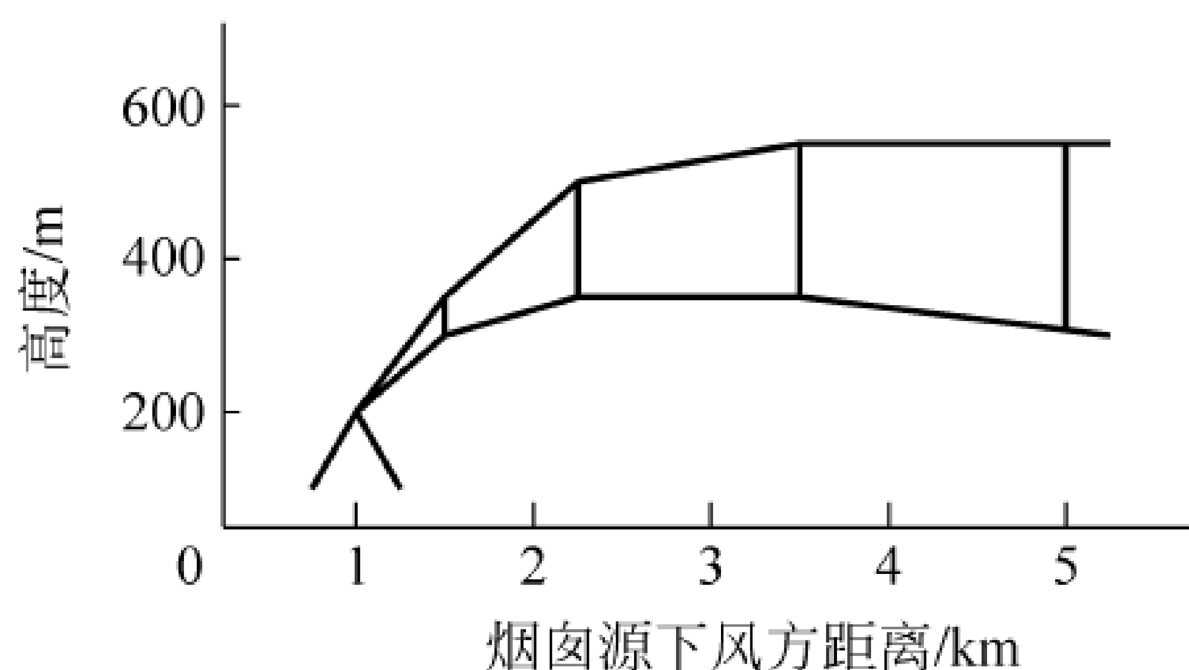



图 10-4-4　激光雷达探测的烟云

除此之外，烟云相对浓度等的测定也是烟云探测的重要方面。综合考虑便可了解到大气污染物的分布实况以及大气污染的扩散规律了。

(3) 激光探测污染气体浓度

激光探测污染气体浓度是激光探测大气污染的一个重要方面。其原理主要有微分吸收原理和喇曼散射原理两种。

微分吸收原理是利用波长与待测气体分子中心吸收谱线(带)重合的激光光束，在大气介质中传输时受到该污染分子的强烈吸收衰减，于是根据激光被衰减的程度便可以确定待测污染气体的浓度。

大气分子中许多污染气体分子都具有丰富的吸收光谱带。例如，在波长小于 400nm 的紫外波段，以及波长为 400～700nm 的可见光波段，一些由双原子分子和不大的多原子分子组成的污染气体分子，如 $O_3$，$SO_2$ 和 $NO_2$ 等主要污染气体分子，均具有吸收光谱带。由此，根据光谱带被吸收的情况就可以推断污染气体的主要成分了。

而激光探测污染气体浓度的喇曼散射原理，是建立在喇曼(Sir Chandrasekhara Venkata Raman，1888—1970)散射的物理基础上的。喇曼散射具有散射光频率不同于照射光频率的特点，而且喇曼散射频移只取决于散射分子的成分，与照射光频率无关。根据喇曼散射的这一特点，可由大气后向散射光的频移特征鉴别污染气体成分，而由喇曼后向散射光强确定污染气体浓度。

以上各种激光探测大气污染的方法在现实生活中是行之有效的，在环境污染严重的今天发挥着不可替代的作用。而更加充分地利用激光的特性应用到各个领域还有待进一步研究。

**3. 激光通信**

激光光纤通信传输的原理就是光的全反射(见图 10-4-5)。激光通信和无线电通信相比，具有以下一些优点：第一，频带宽，通信容量大。光波的频率是 $10^{13}\sim10^{15}$ Hz，比无线电

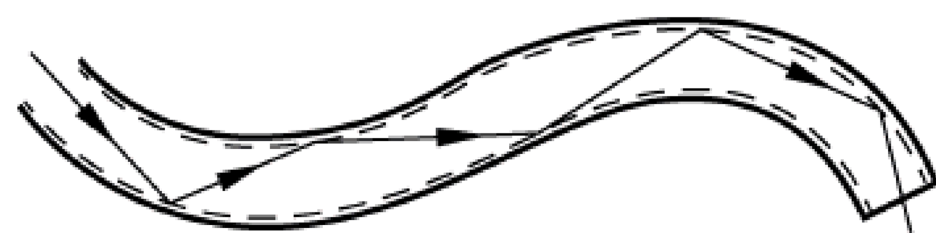


图 10-4-5　激光光纤传输

波高 $10^5$ 倍，比微波也高近 $10^4$ 倍。理论上计算，光通信容量可以传输100亿路电话或1000万套电视节目。因技术上的种种原因，目前实际一对光纤可以通几千路电话。由于一根光纤很细，可以由几十根或更多根光纤组成光缆，因此可以达到很大容量。第二，抗干扰性强。外界的电磁波因与光波频率范围不同，所以不会干扰光通信。第三，设备结构简单，投资少，使用寿命长。第四，采用光导纤维传输激光信号比用有色金属导线传输电信号成本低30%，可以节约大量金属材料。正因为激光通信具有以上这些优点，所以在通信领域起着越来越重要的作用，加快了人类进入信息时代的步伐。

**4. 激光信息处理**

激光除了可以传输信息外，在储存信息方面也具有重要作用，如全息照相、光学计算机、光盘等。以往的光学图像只能记录光的强度，利用激光既可以记录光的强度又可以记录光的相位，即可以记录光的全部信息，照出的图像是立体的，因此称为全息照相。利用激光的相干性，人们正在研究光学计算机。光学计算机与电子计算机相比，具有运算速度快、信息容量大的优点。它的信息处理过程接近于人眼和大脑的观察与识别外界事物的过程。光盘即激光视盘，是利用激光储存信息的“唱片”。它是由两片透明塑料组成，塑料片上涂有一层反射膜。光盘的厚度约1mm，直径为20～30cm，两面都可使用。用光盘播放时，首先由半导体激光器发出激光，一台小型计算机控制激光来扫描，得到以数字形式输入的脉冲信号，再由计算机转换成电视信号传给电视机。用光盘播放的音像质量高，而且光盘中记录的每幅图像都有编号，可以任意选取，非常方便。

**5. 激光医疗**

用激光可以治疗眼疾，如视网膜焊接、虹膜切除、角膜移植、治疗青光眼等。激光在皮肤科、五官科、妇科、肿瘤科、外科等方面也得到了应用。可治疗黑色素皮肤癌、血管肿瘤、焊接骨骼。外科用激光作为手术刀来切割组织，其优点是可以自动止血、伤口基本无菌、伤口愈合快。利用光导纤维传输激光，制成激光内窥镜来检查肠胃的病变。

## 10.5 关于光速的话题

光具有波动性同时具有粒子性，从泊松亮斑的证实到光电效应的发现无不说明了人们对光的认识的不断深入。关于光的本性传奇，一直就是人们研究的热点，其中人们最为关切的是光速。光速不仅仅是光传播的速度，也是信息传递速度的绝对极限。它不仅把时间与空间以一种根本的方式联系在一起，还保证未来不会先于过去发生。而对于光速的研究，随着长久以来科学家们的不断奋斗，也派生出各种各样的说法和结论。

### 10.5.1 光速的定义

光的电磁理论指出，介质中电磁波的传播速度

$$v=\frac{c}{\sqrt{\varepsilon_r\mu_r}}=\frac{1}{\sqrt{\varepsilon\mu}} \tag{10-5-1}$$

其中 $\varepsilon_r$，$\mu_r$ 分别是介质的相对介电常数和相对磁导率。真空中电磁波的速度与光在真空中的速度相同，记为 $c$，有

$$c=\frac{1}{\sqrt{\varepsilon_0\mu_0}} \tag{10-5-2}$$

其中 $\varepsilon_0$，$\mu_0$ 分别是真空的介电常数和磁导率。光在透明介质中的传播速度 $v$ 和真空中的速

度 $c$ 之比为介质的折射率：$n=\dfrac{c}{v}$。由此结论，麦克斯韦认为：光是一种电磁波，于是，$n=\sqrt{\varepsilon_r\mu_r}$。无论是麦克斯韦的电磁波理论，还是爱因斯坦的狭义相对论，都说明光的速度与参考系无关。

### 10.5.2　光速的测量

光速的测定在光学的发展史上具有非常特殊而重要的意义。它不仅推动了光学实验，也打破了光速无限的传统观念；在物理学理论研究的发展里程中，它不仅为粒子说和波动说的争论提供了判定的依据，而且最终推动了爱因斯坦相对论理论的发展。

在光速的问题上物理学界曾经产生过争执，开普勒和笛卡儿都认为光的传播不需要时间，是在瞬时进行的。但伽利略认为光虽然传播得很快，但光速却是可以测定的。1607 年，伽利略进行了最早的测量光速的实验。伽利略的方法是，让两个人分别站在相距一英里的两座山上，每个人拿一个灯，第一个人先举起灯，当第二个人看到第一个人的灯时立即举起自己的灯，从第一个人举起灯到他看到第二个人的灯的时间间隔就是光传播两英里的时间。但由于光速传播的速度实在是太快了，这种方法根本行不通。但伽利略的实验揭开了人类历史上对光速进行研究的序幕。

光以有限但非常高的速度传播这一事实，由丹麦的天文学家罗默（Ole Christensen Rømer，1644—1710）于 1676 年第一次发现。他观察到木星的月亮不是以等时间间隔从木星背后出来，不像如果月亮以不变速度绕木星运动时人们所预料的那样。当地球和木星都绕着太阳公转时，它们之间的距离在变化着。于是，他使用木星的一颗卫星有规律的轨道运动作为计时器，每次这颗卫星被巨大的行星（木星）所掩食，他便记录下一个“滴答”。他发现，从地球上观察，这些滴答的出现并不像预想的那么规律，在一年之中会时而快几分钟，时而慢几分钟（见图 10-5-1、图 10-5-2）。

罗默计算出，这些时延是木星和地球在绕太阳运动时它们之间的距离变化所引起的。通过计算一年里地球、木星及其卫星在轨道上的相对位置，他算出了光穿过宇宙空间的速度。1676 年 9 月，罗默预言预计 11 月 9 日上午 5 点 25 分 45 秒发生的木卫食将推迟 10 分钟。巴黎天文台的科学家们怀着将信将疑的态度，但是观测结果最终证实了罗默的预言。然而，他测量到的木星到地球的距离变化不是非常准确，所以他于 1676 年向法国科学院提交了他的光速结果为每秒 140000 英里。

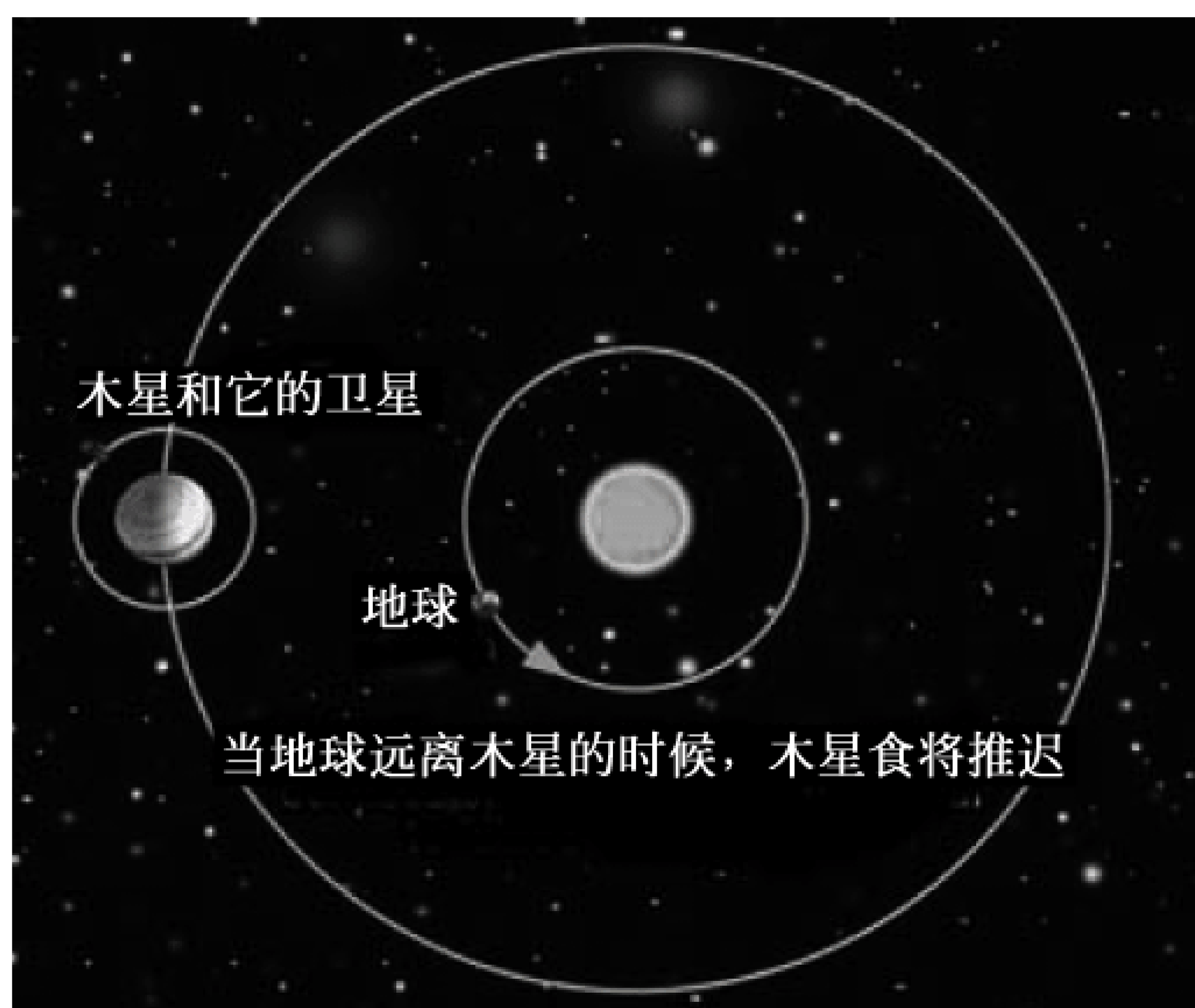



图 10-5-1　地球远离木星，木星食推迟

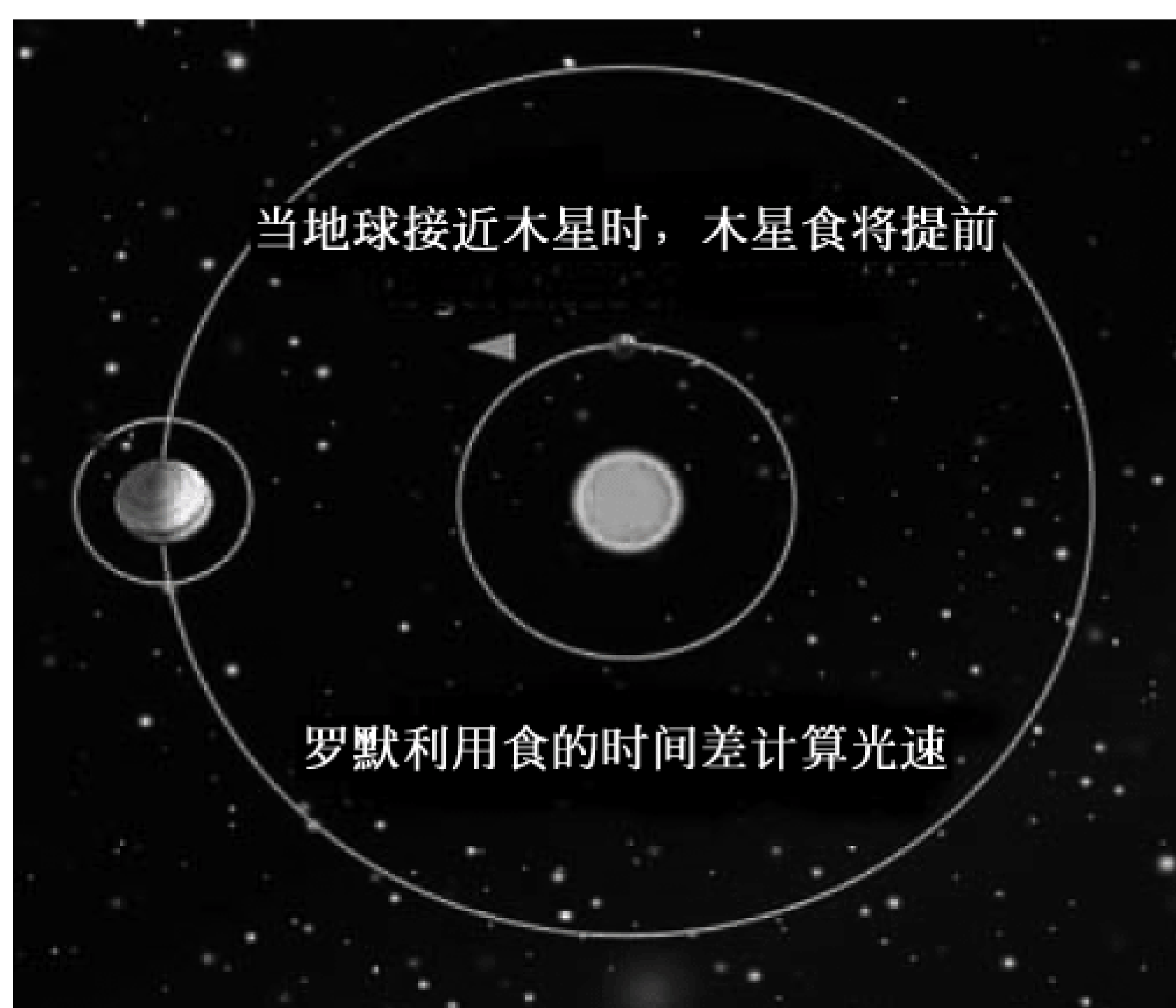



图 10-5-2　地球接近木星，木星食提前

1849 年,法国人菲索(Fiso,1819—1896)第一次在地面上设计实验装置“齿轮法”来测定光速。他的方法的原理与伽利略的相类似,如图 10-5-3 所示,他将一个点光源放在透镜的焦点处,在透镜与光源之间放一个齿轮,在透镜的另一侧较远处依次放置另一个透镜和一个平面镜,平面镜位于第二个透镜的焦点处。点光源发出的光经过齿轮和透镜后变成平行光,平行光经过第二个透镜后又在平面镜上聚于一点,在平面镜上反射后按原路返回。由于齿轮有齿隙和齿,当光通过齿隙时观察者就可以看到返回的光,当光恰好遇到齿时就会被遮住。从开始到返回的光第一次消失的时间就是光往返一次所用的时间,根据齿轮的转速,这个时间不难求出。通过这种方法,菲索测得的光速是 315000km/s。由于齿轮有一定的宽度,用这种方法很难精确地测出光速。

1924—1927 年,美国科学家迈克耳孙综合菲索和傅科(Foucault. Jean-Bernard-Lon,1819—1868)测光速方法的优点,用旋转棱镜法测得了光速。其结果非常接近后来用极其精密的仪器测出的结果。

如图 10-5-4 所示,左边是凹面镜将从右面射过来的光线反射到平面镜,经平面镜反射凹镜再反射变成与原方向相反的光线。右边是八面棱镜将位于上方的光源射到面 1 的光线向左反射到凹镜。反射回来的光线再射到八面棱镜面 3,经面 3 反射后,光线射到望远镜筒,可以被观察者看见。(要指出的是凹镜到八面镜的距离相当远)转动八面镜,这样从面 1 反射的光线从凹镜反射回来时,面 3 已转了过去,在望远镜里就看不到反射光线了。不断地将八面镜转速加快,当快到一定程度,使得光线从凹镜反射回来时,面 2 正好转到了面 3 的位置,望远镜筒里又能看得见这条光线了。光在反射的过程中所花的时间就等于八面棱镜面 2 转到面 3 的位置的时间,只要测出八面棱镜的转速,计算出八面镜转过 1/8 周的时间 $t$,把距离除以这个时间 $t$ 就能得到光在空气里的传播速度了。

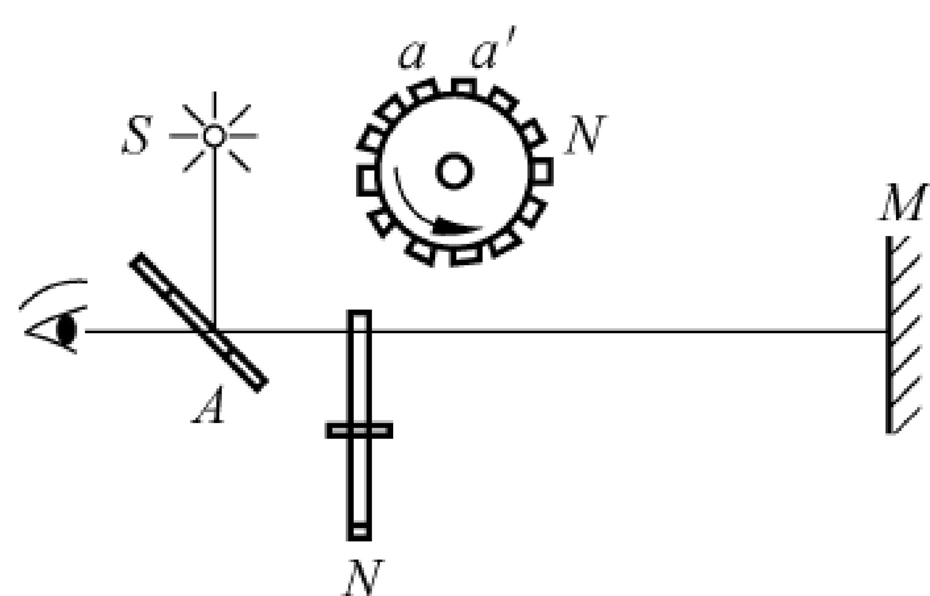



图 10-5-3 菲索齿轮法测光速

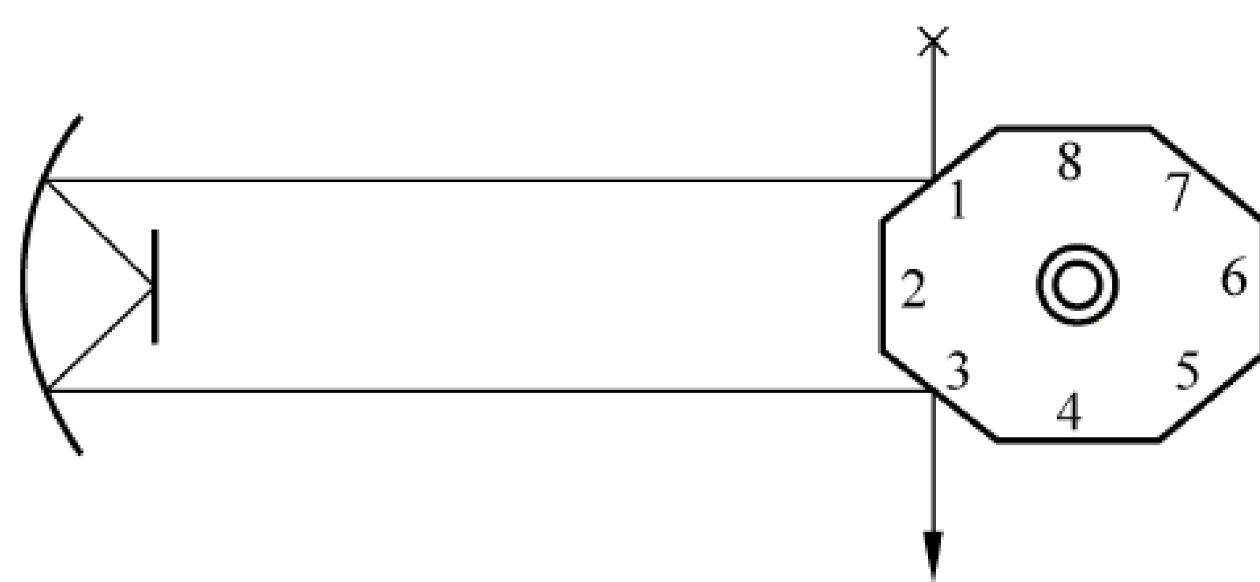



图 10-5-4 旋转棱镜法测光速

1983 年,光速取代了米被选作定义标准,约定为 299792458m/s,数值与当时的米定义一致。秒和光速的定义值表示 1m 从此定义为光在真空中 1/299792458s 内走过的距离。因此自 1983 年以来,不管我们对光速的测量作了多少精确的修正,都不会影响到光速值,却会影响到米的长度表示。

### 10.5.3 光速不变吗

爱因斯坦年轻的时候曾经问自己,如果人运动的速度快到足以跟上光的脚步,光看起来是什么样子的。爱因斯坦知道,麦克斯韦方程组不允许这种结果出现。他得出结论:要么是麦克斯韦的理论不适用于运动中的观察者,要么是相对运动力学需要更改。

爱因斯坦在他 1905 年发表的狭义相对论里基于一个通用原则:相对任何以恒定速度运动的观察者来说,不管这个速度是多少,物理原理及光速都是一样的。爱因斯坦的狭

义相对论使我们对时间和空间的观念发生了革命性的变化，强调了光速在物理学中的根本地位。

想象你在一枚火箭里，与一道激光脉冲一同冲入宇宙空间。地球上的观察者会看到这一脉冲以光速远去。无论你相对于地球运动的速度为多少，譬如，光速的 99%吧，光线仍以光速超越你。看起来似乎很荒谬，但这是真的，使这为真的唯一途径，就是你火箭中的观察者和地球表面的观察者以不同方式衡量时间和空间。

“光速不变原理”涉及光这种物质与其他物质的关系问题。所谓的“光速不变”是一种通俗的、不够严谨的说法。正确的表述应该是：光以其特有的速度运动，它的运动速度与发光体的运动速度无关，也与观测者的运动速度无关。

在相对论中，光速不变是很特殊的，不是我们简单理解的匀速直线运动，而是指在真空中，光没有相对速度。光速在相对论中有两个观点：一是光不具有相对速度，这是相对论的前提；二是光速 $c$ 是运动速度的极限，这是相对论的结论（根据洛伦兹变换）。

但爱因斯坦忽略了一个基本的物理现象：就是多普勒（Doppler Christian Johann，1803—1853）效应。在各处所测得的光速虽同为 $c$，但其频率是不同的，而频率的不同恰恰是运动所引起的，而且与相对运动速度大小成正比。在五维时空坐标中的分析结论知道，这是光与其他物体具有相对运动速度的间接表现。因此对于光、电磁波或微观粒子而言，麦克斯韦电磁理论、相对性原理和伽利略变换同时成立，在不同惯性空间中，光、电磁波或微观粒子还具有洛伦兹变换；宏观物质则是速度 $v$、相对性原理、伽利略变换同时成立，没有洛伦兹变换。进一步分析我们知道，它还有更多的物理内涵。

### 10.5.4 超光速的有趣话题

爱因斯坦在狭义相对论中断言，在真空中没有物质的速度能超过光速。如果超过光速，那么或许出现这样的可能性，即事情的发生早于导致事情发生的原因。

2000 年 7 月，由于英国 *Nature* 杂志发表了一篇关于“超光速”实验的论文，引起了人们对超光速到底是否存在的讨论。其实对在介质中使光脉冲的群速度超过真空中光速 $c$，科学家们早有研究，而 *Nature* 中报道的这个实验就是实现了这种想法。但是这并非是人们想象的那种所谓违反因果律（或者相对论）的超光速，为了说明这个问题，让我们看一看由华人科学家王力军所做的这个实验。

光脉冲是由不同频率、振幅、相位的光波组成的波包，光脉冲的每个成分的速度称为相速度，波包峰的速度称为群速度。在真空中二者是相同的，但是在介质中如我们所知道的存在如下的群速度与介质折射率的关系：

$$V_g = \frac{c}{n_g}, \quad n_g = n + \omega\left(\frac{\mathrm{d}n}{\mathrm{d}\omega}\right) \tag{10-5-3}$$

显然在一定的情况下（如反常色散很强的介质）可以出现负的群速度，此时，光脉冲在介质中传播比真空中的时间短，其差

$$\Delta T = \frac{L}{v} - \frac{L}{c} \tag{10-5-4}$$

达到绝对值足够大时就可以观察到“超光速”现象，即“光脉冲峰值进入介质以前，在另一边已经有脉冲峰出射了”。

那么这种超光速是不是违背因果律呢？我们仔细考查王力军的实验就会发现，出射光脉冲虽然是在入射脉冲峰值进入介质之前出现的，但在这之前入射脉冲的前沿早已进入介

质了,如图 10-5-5 所示。因此出射脉冲可以看作是由入射脉冲前沿与介质相互作用产生的。其实王力军的实验重要意义正在于实现了可观测的负群速度的这一现象,而不是像媒体炒作的那样发现了什么"超光速"。负群速度在这里就不能理解为光的速度了,它也不是能量传输的速度。当然,这一实验本身就说明我们人类对光的认识又前进了一步。对这个实验的解释只凭折射率与群速度的关系这个公式是远远不够的,这其中包含了量子干涉的效应,涉及对光的本质的认识,揭开蒙在"超光速实验"头上的面纱,仍然是科学家们奋斗的目标。

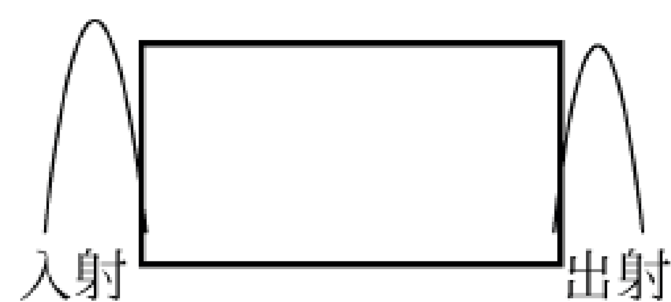
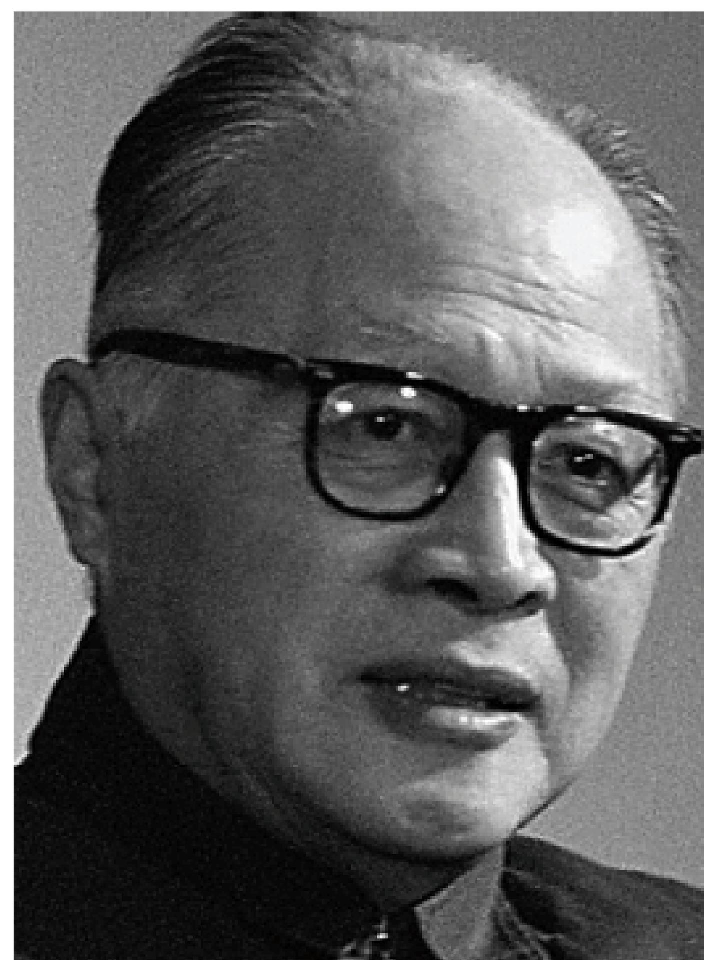


图 10-5-5 王力军超光速实验

## 学习资源

### 一、中国近现代著名科学家(物理学)简介

王淦昌(1907 年 5 月 28 日—1998 年 12 月 10 日),著名核物理学家、中国核科学的奠基人和开拓者之一、中国科学院资深院士、九三学社中央名誉主席、中国共产党优秀党员、原第二机械工业部副部长。江苏常熟支塘镇人。两弹一星元勋,中国实验原子核物理、宇宙射线及基本粒子物理研究的主要奠基人和开拓者,在国际上享有很高的声誉,被誉为"中国核武器之父""中国原子弹之父"。



王淦昌 1929 年毕业于清华大学物理系,1930 年入德国柏林大学,1933 年于德国柏林大学获博士学位。1934 年 4 月回国,1936—1952 年,先后在山东大学、浙江大学任教授,在浙大度过了他一生中最美好的 16 年,并培养了李政道、叶笃正等一大批优秀的科学家。在 70 年科研生涯中,他奋力攀登,取得了多项令世界瞩目的科学成就。

王淦昌 1979 年加入中国共产党,历任第二机械工业部(现中国核工业总公司)九院副院长,二机部副部长兼原子能研究所(现中国原子能科学研究院)所长,中国科学技术协会副主席,中国核学会理事长,九三学社中央参议委员会主任,第三届、四届、五届、六届全国人大常委会委员。1982 年的一天,在江苏省常熟县,一位派出所的同志领着几位外地来客,走街串巷,来到一座木板房前,来客中的一位老人,端详着这修缮过的房屋,激动地连声说:"就是这里,就是这里。"这位老人就是中国著名核物理学家王淦昌。他来探望深深怀念的生他养他的故乡。

陈景润(1933 年 5 月 22 日—1996 年 3 月 19 日),福建福州人。中国著名数学家,厦门大学数学系毕业。1966 年发表《表达偶数为一个素数及一个不超过两个素数的乘积之和》,成为哥德巴赫猜想研究上的里程碑。而他所发表的成果也被称之为陈氏定理。这项工作还使他与王元、潘承洞在 1978 年共同获得中国自然科学奖一等奖。陈景润研究"哥德巴赫猜想"和其他数论问题的成就,至今仍然在世界上遥遥领先。世界级的数学大师、美国学者阿·威尔曾这样称赞他:"陈景润的每一项工作,都好像



是在喜马拉雅山山巅上行走。”1978年和1982年，陈景润两次受到国际数学家大会作45分钟报告的最高规格的邀请。此外，陈景润还在组合数学与现代经济管理、尖端技术和人类密切关系等方面进行了深入的研究和探讨。他先后在国内外报刊上发表了科学论文70余篇，并有《数学趣味谈》《组合数学》等著作，曾获国家自然科学奖一等奖、何梁何利基金奖、华罗庚数学奖等多项奖励。陈景润在国内外都享有很高的声誉，然而他毫不自满，他说：“在科学的道路上我只是翻过了一个小山包，真正高峰还没有攀上去，还要继续努力。”陈景润不善辞令，为人低调，生活简朴，1996年3月19日在骑自行车上班路上，不幸死于车祸。

## 二、诺贝尔物理学奖简介

**1981年诺贝尔物理学奖**。因发展高分辨率电子能谱仪并用以研究光电子能谱和作化学元素的定量分析。获奖者：K. M. 西格巴恩（瑞典）、N. 布洛姆伯根、A. 肖洛（美国）。评价：高分辨率电子能谱仪是通过光电效应，用粒子轰击物体产生电子，然后通过分析这些粒子的动能可以知道它在化合物中的结合能。由于对同一种化合物它的各个电子化合能是一定的，所以可以据此来确定化合物的状态。布洛姆伯根被公认为是非线性光学的奠基人。一般来说一个东西只要是非线性的都是很复杂的。而很多时候激光在介质中传播就是如此。不过布洛姆伯根能创建这项学科可见他的伟大。而肖洛则是通过非线性光学对激光光谱学的研究有重大贡献。另外他还曾是研究激光器的先驱。

**1982年诺贝尔物理学奖**。因建立相变的临界现象理论，即重正化群变换理论。获奖者：K. G. 威尔逊（美国）。评价：先解释一下什么叫临界相变理论。一个物体有一些不同的状态比如固态、液态、气态等。我们知道在不同气压下物体的沸点和凝固点不同就形成了相变曲线。其实不只是固液气三态，物质的很多性质都有类似的地方。比如超导存在临界温度，物体的铁磁性和顺磁性的变化的临界温度（居里点），它们在临界点都有很多类似的性质。这就令人有理由猜想关于任何临界点有相似的性质。这就是临界相变理论。苏联著名物理学家朗道曾经提出过此方面的理论。不过随着实验精度的提高，朗道的理论并不准确。威尔逊将量子力学中的重整化群方法应用到相变理论中，也取得了丰硕的成果。不过距离这一理论的最终完成还有很长的距离。

**1983年诺贝尔物理学奖**。从事星体进化的物理过程的研究。获奖者：S. 昌德拉塞卡、W. A. 福勒（美国）。评价：昌德拉塞卡的最大贡献在于发展了白矮星的理论。提出了昌德拉塞卡极限（即白矮星质量的最大值），成功地解释了恒星的晚期演化，因此对宇宙学做出了重大贡献。贝特提出了太阳和恒星的能源主要来自它们内部的氢碳的聚变反应（1967年诺贝尔奖）。福勒解决了当氢完成聚变后恒星如何反应的问题，而且也解决了各种元素如何产生的问题（通过聚变和超新星爆炸产生）。这次两人的获奖是因在天体物理领域方面的研究而获诺贝尔奖的第8位和第9位科学家。其实说明了人类的两个追求目标。一个是想了解至大的宇宙，另一个是想了解至小的基本粒子。

**1984年诺贝尔物理学奖**。对导致发现弱相互作用的传递者而在场粒子W±和Z0的大型工程做出了决定性贡献。获奖者：C. 鲁比亚（意大利）、S. 范德梅尔（荷兰）。评价：弱电统一理论是20世纪物理学的巅峰成就（1979），量子力学的观点中力是通过粒子的碰撞产生的，而S. L. 格拉肖、S. 温伯格、A. 萨拉姆提出的理论提出弱力的携带粒子为W＋，W－，Z0。不过由于当时的设备不允许所以一直未得到证实。终于欧洲核子研究中心在C. 鲁比亚和S. 范德梅尔的领导下做出了此重要发现。进一步证明了弱电统一理论的正确性。

**1985年诺贝尔物理学奖**。发现量子霍尔效应。获奖者：K. 冯·克里清（德国）。评价：

两个平行平板加一个竖直的磁场，然后有电流通过，就会因为有洛伦兹力的作用，会形成电势差，还会形成一个等效电阻。而量子霍尔效应就是在某些情况下，电阻不是连续变化的而是分立的值。其实这一现象当时已有很多人发现但是却没引起重视。只有冯·克里清作了坚持不懈的研究。当有人问冯·克利青，量子霍尔效应是不是一个偶然的发现？他解释说量子霍尔效应作为一个普遍规律而存在的重大想法是在1980年2月5日凌晨突然闪现出来的，但它是基于长期研究工作之后的一个飞跃。“通过测量大量的不同样品，才第一次认识这样一种特殊的规律，而这种平凡重复的测量简直弄得我们感到乏味，我们反复变化样品，变化载流子浓度，将磁场从零扫描到最大……终于我们发现了这样的特殊规律，所以这一结果的取得是长时间努力工作的结果，这些测量的曲线无时不在我的脑子里盘旋着，反复思考着。”或许坚持也可能是创造力。

**1986年诺贝尔物理学奖**。在电光学领域做了大量基础研究，开发了第一架电子显微镜。获奖者：E. 鲁斯卡(德国)。设计并研制了新型电子显微镜——扫描隧道显微镜。获奖者：G. 比尼格(德国)、H. 罗雷尔(瑞士)。评价：光学仪器的分辨率看似玄乎其实取决于两个东西，一个是用来探测的波长(如光波，微波)，另一个是孔径(通俗地说差不多即是所用的镜子的大小)。所以要提高分辨率只能从两个方面下手。一是像哈勃望远镜那样作个直径几米的镜子(当然也有前面谈到的孔径合成技术，但那对灵敏度有影响)。不过那成本，那重量不是一般地方可以承受的。所以要从波长入手，而实物粒子的波长比电磁波的波长要小很多，而电子是我们最熟悉的实物粒子，所以电子显微镜是科学家们一直想实现的目标。而上述三人或实现或发展了电子显微镜技术，让人类的眼睛的能力提高了上千倍。

**1987年诺贝尔物理学奖**。发现氧化物高温超导体。获奖者：J. G. 贝德诺尔斯(德国)、K. A. 米勒(瑞士)。评价：超导是一项极具诱惑力的技术，自发现起就引起了人们的广泛关注。关于它的理论研究曾连续两次获得诺贝尔奖。但不得不说超导的临界温度提高的并不快。从1911年到1986年，75年间从水银的4.2K提高到铌三锗的23.22K，才提高了19K。但是从1986年起超导出现了飞跃似的发展。以美国、日本和中国为中心的全球性的“超导热”研究，在短短的3个月内，临界温度从33K迅速提高到100K以上。超导也开始进入到实际应用的阶段(磁悬浮列车)。而这个开端就是贝德诺尔斯、K. A. 米勒提出钡镧铜氧化合物的高温超导性。

**1988年诺贝尔物理学奖**。发现$\mu$子型中微子，从而揭示了轻子的内部结构。获奖者：L. 莱德曼、M. 施瓦茨、J. 斯坦伯格(美国)。评价：说到中微就得先谈谈中子。科学家们发现中子在衰变成质子和电子(β衰变)时，能量会出现亏损。物理学上著名的哥本哈根学派鼻祖尼尔斯·玻尔(1922)据此认为，β衰变过程中能量守恒定律失效。不过泡利(1945)却认为这是因为反应是产生了一种新的粒子，不过它的穿透力很强(可以穿过地球)我们无法观察到它。所以中微子假说就是物理学上的一个重要问题。毕竟能量守恒可是物理学的根本所在。1962年，美国哥伦比亚大学的莱德曼、施瓦茨和斯坦博格等人，想到可以用加速器来产生中微子并发现中微子有两种，除β衰变产生的电子中微子外还有一种$\mu$中微子，他们对中微子做出了开创性的研究，甚至为后来建立弱电统一理论奠定了基础。

**1989年诺贝尔物理学奖**。发明了分离振荡场方法及用之于氢微波激射器及其他原子钟；创造了世界上最准确的时间计测方法——原子钟。获奖者：N. F. 拉姆齐(美国)。发展离子捕集技术。获奖者：H. G. 德默尔特、W. 保罗(德国)。评价：1989年三位诺贝尔物理学奖获得者都是在原子物理技术方面做出过杰出贡献的物理学家，他们创造性地发展了精

确的计量方法,大大改进了实验的技术水平,使许多以前无法进行的实验得以实现,并达到前所未有的精确程度。由于他们的工作,科学界有可能对一些基本物理定律进行更深入的检验,从而提高了人类认识物质世界的能力。

**1990 年诺贝尔物理学奖**。因对电子与质子和束缚中子深度非弹性散射进行的先驱性研究以及因此而对粒子物理学中夸克模型的发展起了重要作用。获奖者:J. I. 弗里德曼、H. W. 肯德尔(美国)、R. E. 泰勒(加拿大)。评价:这次实验又是 SLAC(斯坦福直线加速器中心)的杰作,他们发现了粒子的非弹性碰撞(存在动能损失的碰撞),并首次从实验证明了夸克模型的存在。而这个电子-质子深度非弹性散射实验引起了粒子物理学的一系列新进展(包括丁肇中的发现),使粒子物理学进入了“夸克-胶子”时代。

## 思考与练习

1. 黑体辐射是如何引出量子论的?
2. 光电效应与康普顿效应的区别和联系。
3. 解释康普顿效应的物理思想是什么?
4. 概率波与机械波的本质区别是什么?
5. 除了热辐射、光电效应和康普顿效应外,还有哪些实验事实可以说明光具有粒子性?
6. 谈谈对“光速不变原理”的认识。

# 第11章

# 相对论与宇宙学

从17世纪牛顿在伽利略、开普勒工作的基础上建立了完整的经典力学理论，到18世纪卡诺、焦耳等建立了宏观热力学理论，至19世纪麦克斯韦的电磁学理论和波动光学理论，物理学经过三次重大的综合和升华，形成了经典物理学宏伟大厦完善的框架，显示出一种形式上的完整。当时，大多数科学家普遍认为宇宙万物必然按照由精美的数学方程所表达的物理定律永远运动下去，似乎人类对自然的认识已达到完美的境地。然而，也有一些科学家比如开尔文等，针对当时的物理学理论无法解释的一些物理现象和结果表示了担忧。因为归根到底，物理学从本质上来讲，是一门实验科学。在物理学中，每个物理概念的确立，物理原理和定律的发现，都必须有坚实的实验基础。许多关键问题的解决，最后都要诉诸实验。

首先，就麦克斯韦完美电磁理论而言，传送电磁波需要一种媒质，即以太，然而迈克耳孙-莫雷实验却否定了以太的存在，于是，1904年5月，荷兰物理学家洛伦兹决心对此作根本性解释。他发表了《速度小于光速运动系统中的电磁现象》的论文，论文用一切力的作用都与电磁力类似的假设，引入了另一个他自己称之为"运动系统中的本地时间$t'$"，本地时间$t'$不同于不动系统中时间$t$，洛伦兹不仅成功地描述了固体棒长度收缩的物理图像，而且他的理论可以呈现出一种特别简单的形式——也就是熟知的洛伦兹变换。

同一年，法国数学家彭加勒在圣路易国际艺术与科学大会上作的题为《数学物理学的原理》讲演中，已经相当明确地提出了"相对性原理"。根据这一原理，不管是对于一个固定不动的观察者还是对于一个正在均匀平移的观察者来说，各种物理现象的规律应该是相同的，因此，没有也不可能有任何方法来判断我们是否处在匀速运动之中。至于相对性原理的思想，彭加勒早在1895年就在一篇文章中论述过："要证明物质的绝对运动，或者更确切地讲，要证明可称量物质相对于以太的运动是不可能的。"不仅如此，彭加勒在1898年发表的论文《时间之测量》中，还提出了假设光速对所有观察者都是常数的意见。他指出："'光速不变并在所有方向上均相同'是一种公设，没有这一公设，就无法测量光速。"

1905年，爱因斯坦发表了划时代的《论运动物体的电动力学》论文，提出了狭义相对论。爱因斯坦针对麦克斯韦电动力学应用于运动物体时产生的问题进行长时间思考，独树一帜地脱离以太学说，而当时物理学界的几乎所有大师，都相信自然界存在绝对静止的以太。

奥地利物理学家和哲学家马赫在《力学及其发展的批判历史概论》中对牛顿绝对时空和绝对运动的批判，给爱因斯坦以很大启发。在狭义相对论的创立过程中，正是马赫对绝对时间的批判，使爱因斯坦醒悟到时间是可疑的，并从"同时的相对性"这个问题上取得了突破。洛伦兹提出的长度收缩、地方时和洛伦兹变换法则，以及他最先形成的关于物体质量随其运动速度增加而增加的结论，为相对论的诞生提供了实质性的贡献。彭加勒的光速不变假设，则使洛伦兹变换成为理所当然的结论。

然而,我们最后把相对论的建立归功于爱因斯坦,并不是否认洛伦兹和彭加勒的成就,最根本的原因在于爱因斯坦与这两位大师有着截然不同的时空观:洛伦兹的理论是基于以太学说的,在他看来引入的“运动系统的本地时间 $t'$”只不过是一个数学辅助量;而彭加勒一生都坚持以太说,认为爱因斯坦建立的狭义相对论只不过是一种假设,最后会走进死胡同。

爱因斯坦于 1905 年建立的狭义相对论是区别于牛顿时空观的一种新的时空理论,“狭义”表示它只适用于惯性参考系。只有在观察高速运动现象时才能觉察出这个理论同经典物理学对同一物理现象的预言之间的差别。现在,狭义相对论在许多学科中有着广泛的应用,它和量子力学一起,已成为近代物理学的两大基础理论。

1915 年,爱因斯坦从非惯性系入手,研究与认识了等效原理,进而建立了研究引力本质和时空理论的广义相对论。

广义相对论是现代宇宙学的理论基础,现代宇宙学的核心理论是宇宙大爆炸理论。

在第 1 章关于时空观念变革的相对论的讨论中,我们重点在于一些结论的物理意义的阐述,本章将在第 1 章基础上,从相对论和宇宙学的物理思想、结论、成就和人们的认识过程等方面进行比较详细的讨论。

## 11.1 迈克耳孙-莫雷实验

19 世纪中期,麦克斯韦连续发表三篇论文,以严谨而对称的数学形式归纳了电、磁和电磁辐射之间最基本的关系,使人们更深刻地认识了电磁和光的本性统一性,建立了电磁理论,从根本上改变了人类自牛顿时代以来对世界本质的认识。但这一理论需要一种传送电磁波的媒质,即以太。然而,要传播波速如此巨大而且又是横波的光波,以太的物理特性必然自相矛盾,的确令当时的许多物理学家感到非常困惑。有没有以太?以太在哪里?法国物理学家菲涅耳的静止以太学说圆满地解释了光行差现象,似乎以太的确存在。美国实验物理学家迈克耳孙和莫雷于 1881 年和 1887 年两次在太平洋无干扰的岛上实验室,用极其精巧的实验,给出了当时令人沮丧而后来令人振奋的实验结果:他们希望能证实找到静止以太,但实验却是零“结果”,即否定了以太的存在,这就是历史上著名的迈克耳孙-莫雷实验。

### 11.1.1 迈克耳孙进行以太漂移实验的起因

**1. 以太的由来**

以太(ether)的观念可以追溯到古希腊时代。亚里士多德就认为,天体间充满有某种介质。其实,“以太”在那时的意思是青天或者上层大气;在古宇宙学中,其代表着占据天体之间的物质。到了 17 世纪,笛卡儿在 1644 年发表的《哲学原理》中就引用了以太的观念。他认为,“虚空”是不可能存在的,整个宇宙充满着一种特殊的移动物质——以太。它是某种绝对精致的极其稀薄的刚性物质,物质在它里面运动不受任何摩擦阻力;它又是看不见的十分虚玄的东西。由于太阳周围的以太出现漩涡,造成行星围绕太阳运动。1678 年,惠更斯把光振动模拟于声振动,看成是以太中的弹性脉动。但是,后来由于牛顿在物理学界的威望,光的微粒说压倒了波动说,以太理论受到压制。牛顿认为不需要以太,他主张超距作用,倾向于微粒说。1800 年以后,由于托马斯・杨和菲涅耳应用光的波动说成功地解释了光的干涉、衍射和偏振等现象,使光的波动说重新占了上风。在波动说的支持者看来,光既然是

一种波,就一定要靠介质才能传播。光能通过万籁俱寂的虚空,必然有一种介质充满其间,这就是以太。他们把以太看成是无所不在、绝对静止、极其稀薄的刚性介质。例如,1804 年托马斯·杨写道:“光以太充满所有物质之中,很少受到或不受阻力,就像风从草丛中穿过一样。”

19 世纪下半叶,由于麦克斯韦电磁场理论的建立和赫兹实验对电磁波的证实,以太理论又有了新的发展。以太被认为是电磁波的载体,是电磁力的根源,绝对静止的以太成了人们理想的参考系。

**2. 布拉德雷观测光行差**

英国天文学家布拉得雷(James Bradley,1693—1762)从 1725 年起与另一天文学家莫利勒(Mo Lile)对一些恒星(特别是天龙座 $\gamma$ 星)的方位作了一系列的精确观测,目的是想测定恒星的周年视差。他们在做过认真的校正之后,把恒星一年 12 个月的位置折算到天顶,发现都呈圆形轨迹。1727 到 1728 年,布拉德雷继续观测,证明其他一些星体也有这一现象,他百思不得其解。据说,由于他偶然注意到当船改变航向时,船帆上的旗帜飘向不同的方向,才领悟到这一视差是因为地球围绕太阳旋转所致。他写道:“假想 $CA$(见图 11-1-1)是一条光线,垂直地落到直线 $BD$ 上,如果眼睛(指观察者)静止于 $A$ 点,那么不管光线的传播需要一定的时间还是瞬间,物体必然出现在 $AC$ 方向上。但是,如果眼睛(观察者)从 $B$ 向 $A$ 运动,而光的传播又需要时间,光的速度与眼睛(观察者)的速度之比等于 $CA$ 与 $BA$ 之比,则当眼睛(观察者)从 $B$ 运动到 $A$ 时,光从 $C$ 传播到了 $A$……”。

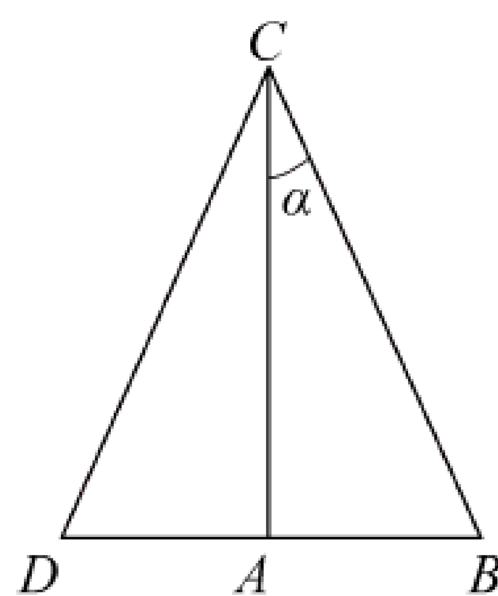


图 11-1-1　布拉德雷的光行差

若用 $\alpha$ 表示$\angle ACB$,$v$ 表示观察者的速度,则有

$$\tan\alpha = \frac{v}{c} \tag{11-1-1}$$

这一关系完全适用于天体的光行差,布拉德雷测到的 $\alpha$ 角为 $40.5''/2\approx 20''$,代入上式得 $c\approx v/\alpha = 3.1\times 10^5\text{km/s}$,其中取 $\alpha = 30\text{km/s}$。

布拉德雷是用微粒说来解释光行差现象的,因为他是微粒说的拥护者。

**3. 阿喇果的实验**

阿喇果是法国天文学家和物理学家。他早年对大气折射有研究,后来对光速问题很感兴趣。他认为测定光的速度是“近代天文学最漂亮的成果”之一。他认为根据牛顿力学中速度叠加的法则,光速将由于发射者和观测者的运动速度不同而略有差别,只是因为布拉雷德的仪器达不到如此高的精确度,才没有显示出这一微小的差别。

于是,阿喇果亲自做了一个实验,在望远镜外,用消色差棱镜加于望远镜视场的半边,再去观测光行差。他考虑,光线经过棱镜与未经过棱镜所受折射不一样,应有不同的光行差角。但是,实际测量的结果却是整个视场的光行差完全相同。其实,这正说明了经典的速度叠加原理不适用于光的传播。而阿喇果并没有想到这一点,却为了拯救微粒说,提出了一个很勉强的假设。他假设星体是以无数种速度发射光的微粒,因为人眼对光有选择性,只能接收某一种特定速度的光,所以看不出有差别。

后来,由于托马斯·杨和菲涅耳的工作,波动说取得了很大进展,阿喇果不久也转向光的波动说。1815 年,他写信给菲涅耳,表示赞同波动说,并且告知 1810 年的实验结果,询问

是不是可以用波动说加以解释。

这件事促使菲涅耳在1818年提出了以太“部分拽引假说”。

**4. 菲涅耳的部分拽引假说**

对于阿喇果的人为选择性的解释,菲涅耳认为很难令人信服。他在给阿喇果的信中,指出这种解释不可取。为了使这两个实验的结果能够协调,他提出了所谓的部分拽引假说,即在透明的物体中,以太可以部分地被运动物体拖曳。他采纳了托马斯·杨的建议,假设透明物体的折射率决定于以太的浓度,并且进一步假设以太的密度与折射率的平方成正比。令 $\rho$ 及 $\rho_1$ 分别表示真空中和透明介质中以太的密度,则

$$\rho/\rho_1=1/n^2=c_1^2/c^2,\quad \text{或}\ \rho_1=n^2\rho \tag{11-1-2}$$

式(11-1-2)中 $c$ 为真空中的光速,$c_1$ 为透明物体中的光速,$n$ 为透明物体的折射率。由此可见,透明物体中的以太比真空中的以太密。菲涅耳进一步假设,真空中的以太是绝对静止的,当透明物体运动时,物体只能带动多于真空的那一部分以太,并设透明物体运动速度 $v$ 与光的传播方向一致,推算出在透明物体中,光的绝对速度等于

$$\frac{c}{n}+\left(1-\frac{1}{n^2}\right)v \tag{11-1-3}$$

其中 $1-\dfrac{1}{n^2}=k$,叫作菲涅耳部分拽引系数。如果 $n=1$,有 $k=0$,则以太完全不受拖曳。这一结果即解释了光行差现象,也解释了阿喇果的实验。

1846年,英国物理学家斯托克斯认为把以太分成不动和可动的两部分,似乎过于牵强,不如假设物体能够完全拖曳,而物体表面不远处的空间中,以太完全静止,中间则是一个逐步减慢的区域。这就叫斯托克斯完全拽引假说。这一假说同样也可以解释阿喇果的实验,而且似乎比菲涅耳的部分拽引假说还更近情理。但是由于不久菲索在1851年所做的流水中的光速实验,证明了菲涅耳的公式,所以斯托克斯的假说不大受人重视。

**5. 菲索的流水实验**

菲索在1851年做的流水实验,设计得很巧妙,如图11-1-2所示。从光源 $S$ 发出的光,经过半透的镀银面 $G$ 反射后,分别通过狭缝 $S_1$、$S_2$ 分成两束光,一束顺水流方向,一束逆水流方向,均经反射镜 $M$ 反射后,在 $S'$ 处发生干涉,观察其干涉条纹,可以检定由于受流水拽引形成的光程差。

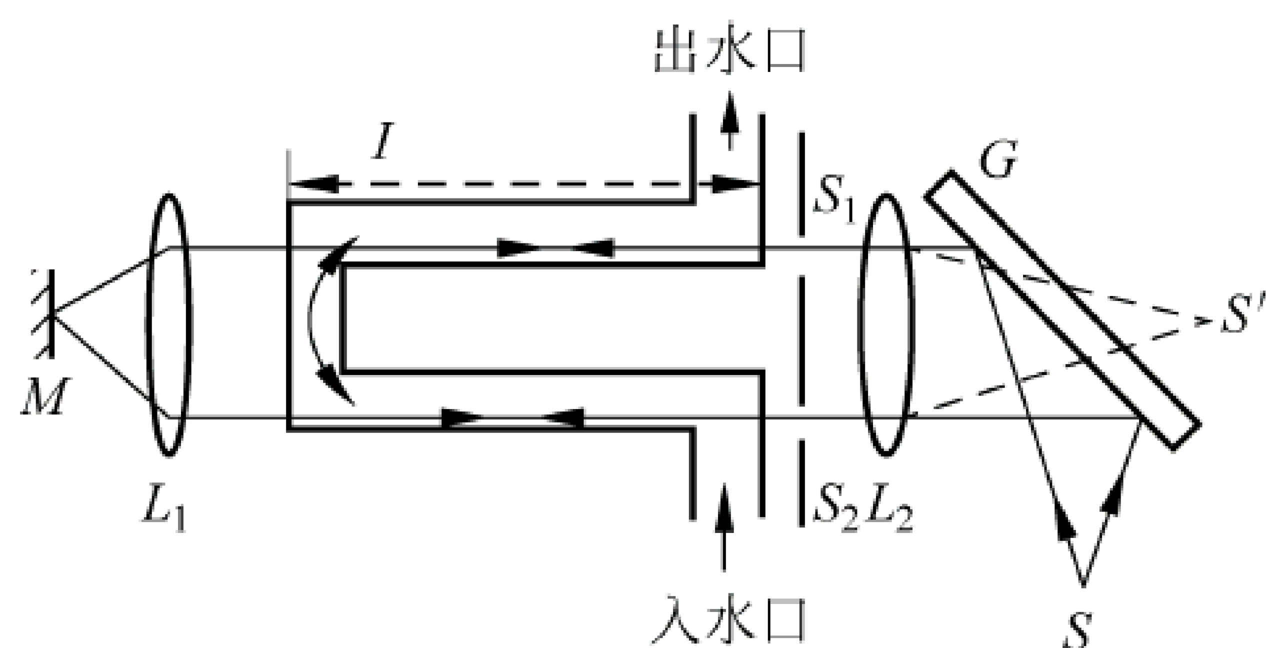


图11-1-2 菲索的流水实验

假如以太不被流水拽引,两束光在流水中的速度一样,所得的干涉条纹将不受流水影响。如果以太被流水拽引,干涉条纹就会受流水影响。

菲索用的是黄光,波长为 $5.26\times10^{-7}$m,水管长度 $l=12$m,水的折射率 $n=1.33$,流速 $v=7.059$m/s,观察到条纹平均移动 $\delta=0.23$ 条。用 $1-\dfrac{1}{n^2}=k$ 计算,预期值为 $\delta'=0.2022$ 条,于是菲索做出结论:“相反两个方向的光速度接近相等”,即以太似乎没有被流水拽引。

**6. 菲涅耳假说进一步被证实**

1868年霍克曾做过类似于菲索的实验,如图11-1-3所示。他也是用半透射的镀银面 $M$ 将光源 $S$ 发出的单色光分成两部分,1路经 $M$ 反射,又经 $M_3$、$L$、$M_2$、$M_1$ 再回到 $M$;2

路经 $M$ 透射，又经 $M_1$、$M_2$、$L$、$M_3$ 也回到 $M$。两路光汇合后，在望远镜中产生干涉条纹。然后整个仪器旋转 180°。实验证明，仪器转 180°，光的干涉条纹没有变化。经过计算，霍克证明了拽引系数不是 0，也不是 1，而正是菲涅耳的结论 $1-\dfrac{1}{n^2}=k$。

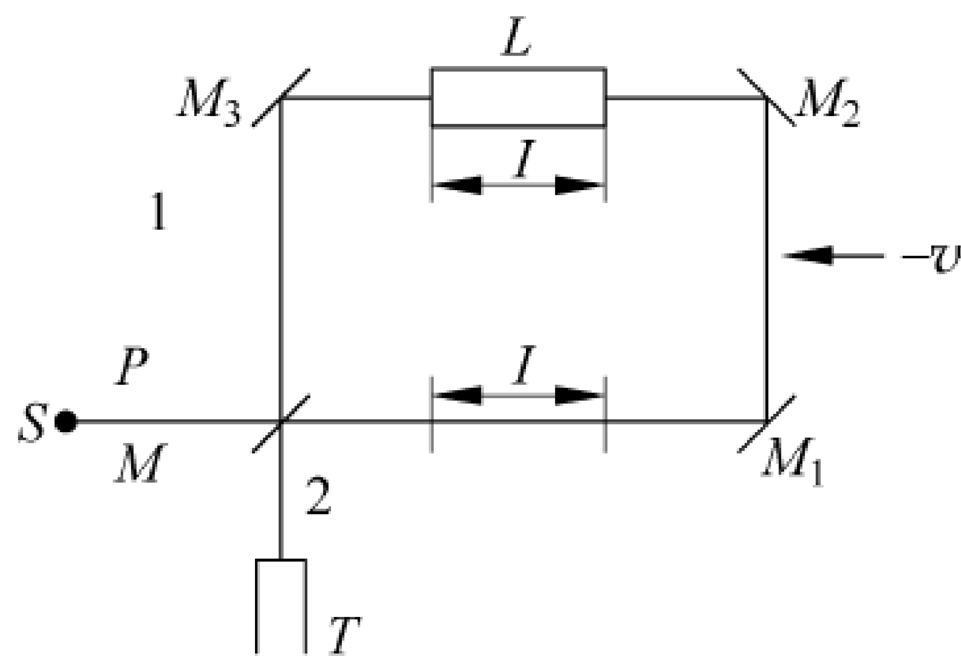


图 11-1-3 霍克证明拽引系数的实验图

菲涅耳的部分拽引假说一而再地得到实验证实，使它成了 19 世纪以太学说的重要组成部分。但是，由它得到的另一结论，却始终没有动静，就是 $n=1$ 时，拽引系数为 0，以太静止不动。于是，以太和运动物体之间就有相对运动。从运动物体上看，以太有漂移速度。那么，地球沿轨道运动，将会在相反方向出现以太风。如果这样，就给人们提供了一种可能的途径：通过测量以太相对于地球的漂移速度，来证实以太的存在和探求以太的性质。

**7. 麦克斯韦的建议**

直到 1879 年，所有试图探测以太漂移速度的实验都没有成功。麦克斯韦很关注这件事，并对此作了论述。他在为《大英百科全书》撰写的“以太”条目中写道：“如果可以在地面上从光由一站到另一站所经的时间测得光速，我们就可以比较相反方向所测的速度，来确定以太相对于地球的速度。然而，实际上地面上测光速的所有方法都取决于两站之间的来回旅程所增加的时间。以太的相对速度等于地球轨道速度，由于这一速度增加的时间仅为整个传播时间得亿万分之一，所以的确难以观察。”

## 11.1.2 迈克耳孙-莫雷实验

**1. 迈克耳孙最初的以太漂移实验**

1880 年，迈克耳孙在柏林大学开始筹划用干涉方法进行以太漂移实验。

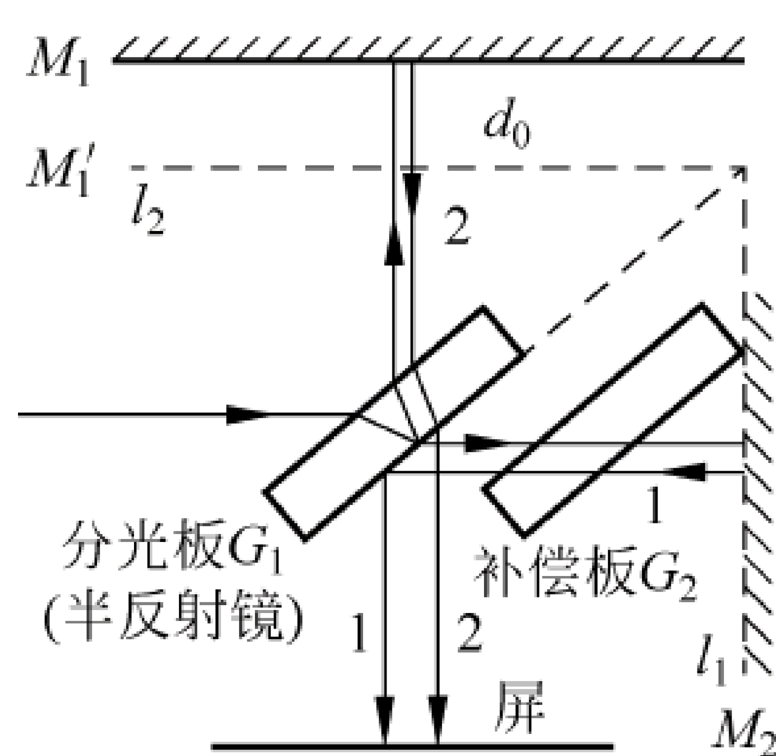


图 11-1-4 迈克耳孙干涉仪的光路图

迈克耳孙把两束反射光的干涉，改为利用一束反射光和一束折射光进行干涉，这样就可以使光的信道在更为广阔的范围内展开。为了检验地球相对于以太的漂移速度，这两束光应该互相垂直，如图 1-11-4 所示。

迈克耳孙巧妙地利用半透明的镀银玻璃板，以 45°角取向将入射光分成互相垂直的两束，经另外两片镜面反射，再会聚到屏(望远镜)中产生干涉条纹。设以太的漂移速度为 $v$，$v$ 与 $l_1$ 臂平行，与 $l_2$ 臂垂直，则光束 1 从 $G_1$ 经 $M_2$ 回到 $G_1$ 的过程所需时间为 $t_1$，有

$$t_1=\frac{l_1}{c-v}+\frac{l_1}{c+v}=\frac{2l_1}{c}\left(\frac{1}{1-v^2/c^2}\right) \tag{11-1-4}$$

设光束 2 从 $G_1$ 经 $M_1$ 再回到 $G_1$ 所需的时间为 $t_2$。由于以太正以速度 $v$ 垂直于光路 $l_2$ 漂移，根据速度合成法则可以推得合速度应为$\sqrt{c^2-v^2}$，所以

$$t_2=\frac{2l_2}{\sqrt{c^2-v^2}} \tag{11-1-5}$$

两束光到达望远镜的时间差为

$$\Delta t = t_1 - t_2 = \frac{2l_1/c}{1 - v^2/c} - \frac{2l_2/c}{\sqrt{1-(v^2/c^2)}}$$

$$\approx \frac{2l_1}{c}\left(1+\frac{v^2}{c^2}\right) - \frac{2l_2}{c}\left(1+\frac{1}{2}\frac{v^2}{c^2}\right) \tag{11-1-6}$$

如将整个仪器转 90°,时间差变为

$$\Delta t' \approx \frac{2l_1}{c}\left(1+\frac{v^2}{2c^2}\right) - \frac{2l_2}{c}\left(1+\frac{v^2}{c^2}\right) \tag{11-1-7}$$

时间差的改变将导致移动 $\delta$ 个干涉条纹,可以求得

$$\delta = \frac{l_1 + l_2}{\lambda c^2} v^2 \tag{11-1-8}$$

如果 $l_1 = l_2 = l$,则

$$\delta = \frac{2(v/c)^2}{\lambda / l} \tag{11-1-9}$$

迈克耳孙根据已知数据:地球的公转轨道速度为 30km/s,$v/c = 10^{-4}$,$\lambda \approx 6\times10^{-7}$m 和 $l = 1.2$m,估算出 $\delta = 0.04$ 个条纹。

一开始,迈克耳孙在柏林大学做实验,因震动干扰太大,无法进行观测,于是改到波茨坦天文台的地下室。实验在 1881 年完成。可是,出乎迈克耳孙的意料,他看到的条纹移动远比预期值小,约为 0.004 到 0.005 条纹,完全在实验误差范围内,而且所得结果与地球运动没有固定的相位关系。于是,迈克耳孙大胆做出假设:“结果只能解释为干涉条纹没有位移。可见,静止以太的假说是不对的。”

但是,这与当时普遍接受的光行差的解释直接矛盾,而且迈克耳孙认为自己的实验精度还不够高,资料计算也有误(估计的效果比实际大两倍),所以,迈克耳孙自己也不满意,认为这次实验不成功。他本人以后也转到精密测定光速的研究去了,实验改进工作就这样搁置起来。

**2. 迈克耳孙-莫雷实验**

1884 年,威廉·汤姆孙(William Thomson,即开尔文勋爵:Lord Kelvin,1824—1907)访问美国并进行讲学,迈克耳孙有机会聆听了这次演讲,并会见了与汤姆孙一起访美的瑞利,就 1881 年的实验交换了意见。他们给了迈克耳孙许多劝告和鼓励,这给了他很大的勇气,激起了他重做实验的兴趣,并下决心与著名的化学家莫雷合作,继续作测量以太漂移的实验。为了提高实验的精度,他们改进了实验装置。为了维持稳定,减小震动的影响,他们把干涉仪安装在很重的石板上,并使石板悬浮在水银面上,可以平稳地绕中心轴转动。为了尽可能增加光程,尽量使干涉仪的臂长增大,他们还在石板上安装了多面反射镜,使两束光来回往返八次,有小光程达 11m。装置如图 1-11-5 所示。

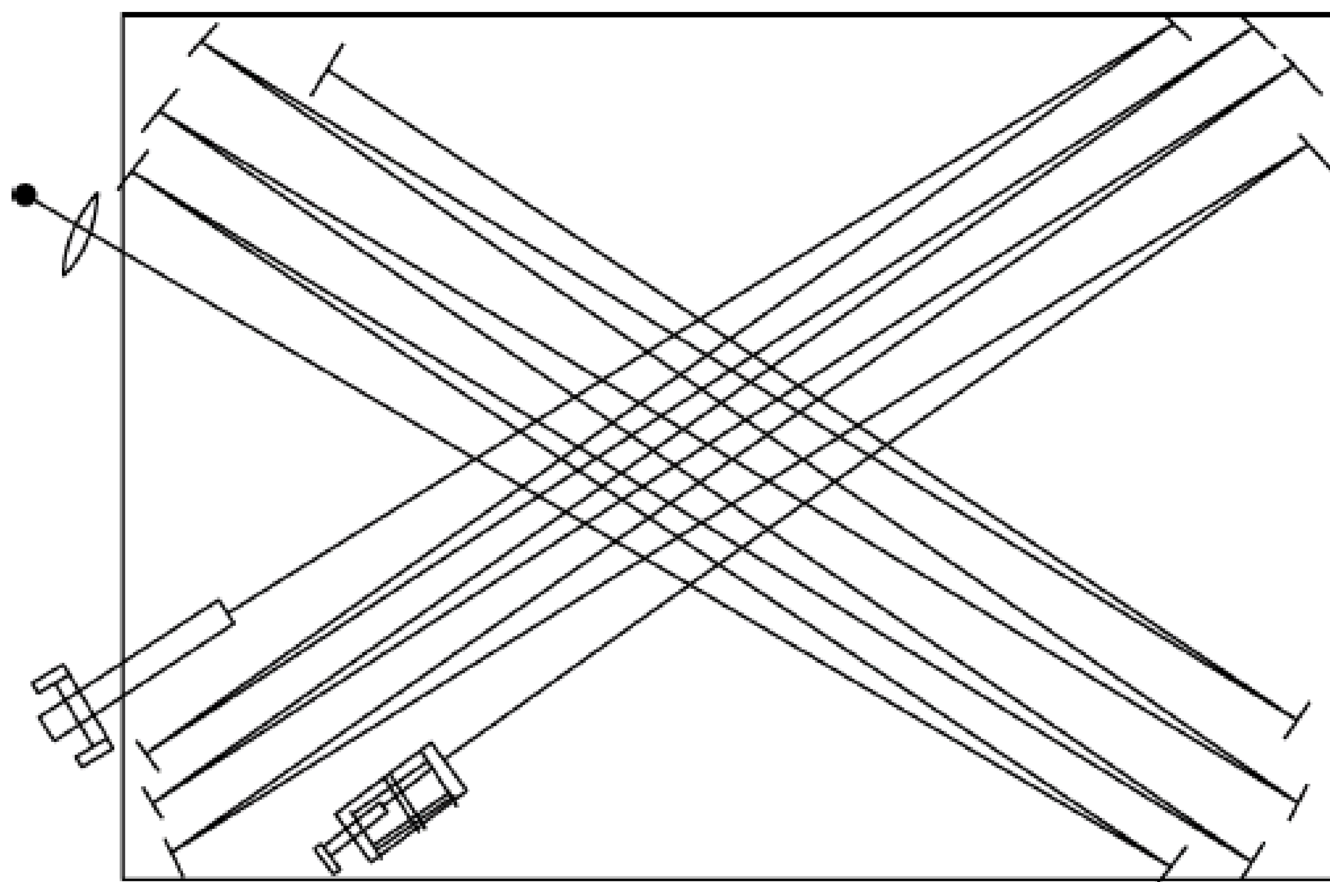


图 11-1-5　迈克耳孙-莫雷实验装置图

他们在白天和夜晚(考虑到地球自转的因素)以及一年的各个季节(考虑地球绕太阳公转的因素)都进行观察。该仪器的精度很高,如果有小到百分之一条纹的移动,该

仪器都可以测出来。

1887 年 7 月,他们在克利夫兰州的阿德尔伯特学院的主楼底层共讨论了四天时间,正式进行实验。他们在论文中描述道:"下表给出六个读数的平均值。表 4 是中午作的观测,表 5 是下午六点钟左右作的观测。读数是鼓轮上的分格值。(干涉)条纹宽度在 40～60 分格之间,平均值近于 50,故一分格相当于 0.02 波长。……观察结果的曲线可以肯定:即使由于地球与光以太之间的相对速度会使条纹产生任何位移,这位移不可能大于条纹间距的 0.01。"

根据理论推算,预期的条纹位移应该为 0.4 个条纹,"实际观测所得的位移,肯定小于这一预期值的二十分之一。由于这位移与速度平方成正比,地球和以太间的相对速度也许小于地球轨道速度的六分之一,肯定小于四分之一。"

迈克耳孙和莫雷对上述的实验结果非常失望,以至于连原来打算在不同季节继续做实验的想法也打消了。

### 11.1.3 迈克耳孙-莫雷实验的意义

迈克耳孙-莫雷实验的零结果对菲涅耳部分拽引假说是一个致命的打击。看来斯托克斯的完全拽引假说比菲涅耳部分曳假说更符合实际。许多人试图用实验来比较这两种假说,其中洛奇在 1892 年做的钢盘实验和迈克耳孙在 1897 年做的不同高度的以太漂移实验最有代表性。然而,这两个实验的结果又是相互矛盾的。

洛伦兹在 1892 年分别提出的收缩假说。这是一个折中方案,既能保留菲涅耳部分拽引假说,又能解释迈克耳孙-莫雷实验的零结果。该假说的基本内容是:物体沿其穿行于以太的方向有收缩,收缩因子为$\frac{1}{\sqrt{1-v^2/c^2}}$,速度 $v$ 越大,收缩因子也越大。这个煞费苦心的假说已经越来越接近正确答案了,但仍未能跳出绝对时空观的牛顿力学的框架。物理学理论面临着一番彻底的变革。

迈克耳孙由于创制了一种新的精度极高的测量仪器,在"精密光学仪器和利用这些仪器进行光谱学的基本量度"方面的研究工作而获得 1907 年的诺贝尔物理学奖。迈克耳孙-莫雷实验的历史意义没有得到当时学术权威的承认。但是,实验结果却动摇了 19 世纪占统治地位的以太假说,为爱因斯坦创立狭义相对论铺平了道路。

迈克耳孙-莫雷实验和爱因斯坦建立狭义相对论到底有没有直接关系,目前有两种不同的说法。一种说法是,爱因斯坦在建立狭义相对论时并不知道迈克耳孙的实验,他是通过洛伦兹的著作知道这个实验的,在 1905 年以后这个实验才引起了他的注意,他说:"否则,我会在我的论文中提到它。"但是,在《我是怎样创立相对论的》一文中又说:"还在学生时代,我就知道迈克耳孙实验的奇怪结果,我很快得出结论,如果我们承认迈克耳孙的零结果是事实,那么地球相对以太运动的想法就是错的,这是引导我走向狭义相对论的最早想法。"

这两个说法前后矛盾,但都不是出自爱因斯坦亲笔。《我是怎样创立相对论的》一文也是他人根据爱因斯坦的即席演说的记录写成的,到底哪个是事实,很难断定。如果从爱因斯坦在 1905 年及以后发表的五篇划时代的论文中没有列出参考文献来看,似乎前一种说法比较可信。但不管怎样,迈克耳孙-莫雷实验对这一理论的证实确实起了决定性的作用。该实验确认了真空中的光速在不同惯性参考系中和各个不同方向上都是相同的,光速是个不变量。这一点对狭义相对论和广义相对论的确证都是非常重要的。这也是为什么许多实验工作者在以后很长时间内仍继续做这个实验的原因。

## 11.2 狭义相对论

时间、空间是什么,别人在很小的时候就搞清楚了,我智力发展迟缓,长大了还没有搞清楚,于是一直揣摩这个问题,结果也就比别人钻研得更深一些。

——爱因斯坦

光以太学说与牛顿力学所引出的“绝对空间”理论紧密相连。牛顿在1687年出版的《自然哲学的数学原理》一书中为我们给出时间和空间的第一个数学模型。牛顿认为:“绝对空间由于它的本性以及同它的外界事物无关,它永远是同一的和不动的。”

既然以太这种东西无处不在,又绝对静止,不受任何东西的干扰,于是,牛顿认为可以把以太看作是绝对的参考体系,它决定了世界上一切运动的永恒的绝对状态。也就是说,我们可以随时确定任何东西的位置,只要标出它在绝对静止的以太中的位置就可以了。

既然存在绝对的位置和状态,牛顿进而认为,也必然存在着“绝对时间”。他说:“绝对的、真正的数学时间自身在流逝着,它的本性是均匀的。它的流逝同任何事物无关。”这种观点认为,时间在均匀地流逝,并且想象在宇宙中有一种“标准钟”,人们可以从放置在任意地方的这种时钟上读出“绝对时间”。任何时间都是绝对的,可以和某个标准时间相比较的。

### 11.2.1 经典物理学面临的挑战

虽然绝对时间和绝对空间是牛顿力学的根基。然而,牛顿的绝对时间和绝对空间有明显的缺陷:既然绝对时间和绝对空间同任何外界事物没有关系,那么怎样才能知道绝对时空的存在?换句话说,这个没有边界的和没有过去未来的绝对时空是怎么存在的呢?因为按照牛顿的观点,它们是无须测量的,甚至根本是无法测量的。这个问题,牛顿也没办法回答。

由于以太是一种假想的“物质”,人们为了解释光和电磁现象,只能根据光和电磁的行为,推测以太的特性,却无法直接用实验证明以太的存在。根据“以太”理论应得出,光线传播速度相对于“以太”应是一个定值,因此,如果你沿与光线传播相同的方向行进,你所测量到的光速应比你在静止时测量到的光速低;反之,如果你沿与光线传播相反的方向行进,你所测量到的光速应比你在静止时测量到的光速高。但是,一系列实验都没有找到造成光速差别的证据。

在这些实验当中,迈克耳孙和莫雷1887年在美国俄亥俄州克里夫兰的凯斯研究所所完成的测量,是最准确细致的。他们对比两束成直角的光线的传播速度,由于围着自转轴的转动和绕太阳的公转,根据推理,地球应穿行在“以太”中,因此上述成直角的两束光线应因地球的运动而测量到不同的速度,莫雷发现,无论是昼夜或冬夏都未引起两束光线光速的不同。不论你是否运动,光线看起来总是以相对于你同样的速度传播。

1900年元旦,英国著名物理学家开尔文勋爵在一篇展望20世纪物理学的新年献辞中说道:“在已经建成的科学大厦中,后辈物理学家只要做一些零碎的修补工作就行了。”“但是在物理学晴朗天空的远处,还有两朵小小的乌云。”指的是当时物理学无法理解的两个实验:热辐射实验和迈克耳孙-莫雷实验。这位科学大师也许预感到“这两朵小乌云”会带来物理学的天翻地覆。其实,这是物理学的一场伟大的革命。实验上一系列新发现,跟经典物理学的理论体系产生了尖锐的矛盾,暴露了经典物理理论中的隐患,指出了经典物理学的局限性。物理学只有从观念上、从基本假设上,以及从理论体系上来一番彻底的变革,才能适

应新的形势。

### 11.2.2 狭义相对论的建立

1905年6月,爱因斯坦在《论动体的电动力学》一文中,第一次提出了崭新的时间空间理论,一举解决了光速的不变性与速度合成法则之间的矛盾以及电磁理论中的不对称等问题。爱因斯坦把这个理论称为相对论理论,简称相对论,后来又叫狭义相对论。狭义相对论是爱因斯坦伟大一生中取得的最有划时代意义的重大成果,是他在前人的基础上经过长期的酝酿和探索才取得的(表11-2-1)。

**表11-2-1 走在爱因斯坦前面的人**

| 人　名 | 贡　献 |
| --- | --- |
| 麦克斯韦(James Clerk Maxwell,1831—1879) | 创建电磁场理论 |
| 赫兹(Heinrich Hertz,1857—1894) | 修改麦克斯韦方程 |
| 弗格特(Woldemar Voigt,1850—1919) | 1887年提出弗格特变换,与洛伦兹变换相似 |
| 拉莫尔(Larmor,1857—1942) | 拉莫尔进动和拉莫尔变换 |
| 菲茨杰拉德(George Francis Fitzgerald,1851—1901) | 独立提出洛伦兹变换 |
| 洛伦兹(Hendrik Antoon,1853—1928) | 提出电子论和洛伦兹变换 |
| 彭加勒(Jules Henri Poincare,1854—1912) | 提出相对性原理 |

到了20世纪初,大量实验和理论研究,为狭义相对论的创建已经准备了必要的条件。但是经典理论的烙印太深刻了,他们无法摆脱绝对时空观的束缚。他们为狭义相对论的创立准备了条件,却没能够创立狭义相对论。

**1. 爱因斯坦关于狭义相对论的思考**

相对论的创立并非一蹴而就。正如爱因斯坦1952年3月11日写给泽利希(Seelich)的信中所说:"在狭义相对论的思想概念和相应的结果发表之间大约经过了五六周时间。但是,如果把这认为是诞生的日期,也许很难说是正确的,因为早在几年前,论据和建筑材料就已经准备好了,虽然那时我还没有下根本的决心。"

的确,有关狭义相对论的思想概念在爱因斯坦的头脑里足足酝酿了10年。

根据爱因斯坦1946年写的《自述》和1922年在日本京都大学的讲演:《我是怎样创立狭义相对论的?》以及其他资料,我们可以追溯他走过的道路。早在16岁(1895年)时,爱因斯坦就开始思考这样一个问题:"如果我以速度$c$(真空中的光速)追随光线运动,我应当看到这样一条光线,就好像一个在空中振荡着而停滞不前的电磁场。可是无论是依据经验,还是按照麦克斯韦方程,看来都不会有这样的事情。从一开始,在我直觉地看来就很清楚,从这样一个观察者的观点来判断,一切都应当像一个相对于地球是静止的观察者所看到的那样按照同样的一些定律进行。因为,第一个观察者怎么会知道或者能够判断他是处在均匀的快速运动状态中呢?"这些悖论使爱因斯坦感到"惊奇",他为此沉思了10年。

实际上,这个悖论已包含了狭义相对论的萌芽。爱因斯坦对这个问题的思考,经历了很长的过程。他回忆说:"最初当我有这个想法时,我并不怀疑以太的存在,不怀疑地球相对以太的运动。"甚至他还设想用热电偶做一个实验,比较沿不同方向的两束光线所放出的热量。

不久爱因斯坦得知迈克耳孙-莫雷实验的零结果。他由此认识到,地球相对于以太的运动是不能用任何仪器测量的。他继续回忆说:"如果承认迈克耳孙的零结果是事实,那么地

球相对于以太运动的想法就是错的，这是引导我走向狭义相对论的第一步。”

后来，爱因斯坦读到了洛伦兹1895年的论文，对洛伦兹方程发生了兴趣。他很欣赏洛伦兹方程不但适用于真空中的参考系，而且适用于运动物体的参考系。他试图用洛伦兹方程讨论斐索的流水中光速实验。当时他坚信麦克斯韦和洛伦兹电动力学方程是正确的，但是进一步推算，发现要保持这些方程对动体参考系同样有效，必然导致光速不变性的概念，而光速的不变性明显地与力学的速度合成法则相抵触。

爱因斯坦喜欢阅读哲学著作，并从哲学中吸收思想营养，他相信世界的统一性和逻辑的一致性。相对性原理已经在力学中被广泛证明，但在电动力学中却无法成立，对于物理学这两个理论体系在逻辑上的不一致，爱因斯坦提出了怀疑。他认为，相对论原理应该普遍成立，因此电磁理论对于各个惯性系应该具有同样的形式，但在这里出现了光速的问题。光速是不变的量还是可变的量，成为相对性原理是否普遍成立的首要问题。当时的物理学家一般都相信以太，也就是相信存在着绝对参考系，这是受到牛顿的绝对空间概念的影响。但是，19世纪末，马赫在所著的《发展中的力学》中，批判了牛顿的绝对时空观。

**2. 狭义相对论的遭遇和实验检验**

由于人们的思想长期受到传统观念的束缚，一时难以接受崭新的时空观，爱因斯坦的论文发表后，相当一段时间受到冷遇，被人们怀疑甚至遭到反对。迈克耳孙至死（1931年）还念念不忘“可爱的以太”，认为相对论是一个怪物。J. J. 汤姆孙在1909年宣称：“以太并不是思辨哲学家异想天开的创造，对我们来说，就像我们呼吸空气一样不可缺少。”1911年美国科学协会主席马吉（M. F. Magie）说：“我相信，现在没有任何一个活着的人真的会断言，他能够想象出时间是速度的函数。”被爱因斯坦誉为相对论先驱的马赫，竟声明自己与相对论没有关系，“不承认相对论”。有一位科学史家叫惠特克（S. E. Whittaker）在写相对论的历史时，竟把相对论的创始人归于彭加勒和洛伦兹，认为爱因斯坦只是对彭加勒和洛伦兹的相对论加了一些补充。就连1922年爱因斯坦获诺贝尔物理学奖也不是由于他建立了相对论。

普朗克和闵可夫斯基（Hermann Minkowski，1864—1909）可以说是支持相对论的代表。正是普朗克，当时作为《物理学年鉴》的主编，认识到爱因斯坦所投论文的价值，及时地予以发表。闵可夫斯基本是爱因斯坦的老师，1908年发表《空间与时间》一文，把空时-时间合并成四维空间，重新处理了相对论的基本方程，把洛伦兹变换看成是空间-时间四维坐标的变换。这样就可以使相对论的规律以更加简洁的形式表达出来。

关于狭义相对论受人们怀疑和反对的情况，可以举电磁质量的实验检验来做些说明。狭义相对论有一重要结果，就是预言电子质量会随运动速度增长。从经典电磁理论出发也可以得到类似的结论。1901年，实验物理学家考夫曼（W. Kaufmann，1871—1947）用β射线的高速电子流进行实验，证实电子的质量确实是随速度变化的。洛伦兹到1904年则根据收缩假说也推出了电子质量公式。后来证明洛伦兹公式与狭义相对论的结果一致。1906年，考夫曼宣布，他的量度结果证实了经典电磁理论家亚伯拉罕（M. Abraham）推导出的电子的质量理论公式，而“与洛伦兹-爱因斯坦的基本假定不相容”。这件事一度竟成了否定相对论的重要依据。在这一事实面前，洛伦兹失望了，他表示，“不幸我的电子变形假说与考夫曼的新结果矛盾，我只好放弃它了。”

然而，爱因斯坦却持另一种态度，他在1907年写文章表示，相信狭义相对论是经得起考验的，在他看来那些理论在很大的程度上是由于偶然碰巧与实验结果相符。果然，一年后布

雪勒(A. H. Bucherer,1863—1927)用改进了的方法测电子质量,得到的结果与洛伦兹-爱因斯坦公式符合甚好。以后许多实验都证明,狭义相对论的结果是正确的。可是,观念的改变不是一朝一夕之事。1911 年索尔维会议召开,由于爱因斯坦在固体比热的研究上有一定影响,人们才注意到他在狭义相对论方面的工作。只是到了 1919 年,爱因斯坦的广义相对论得到了日全食观测的证实,他成为公众注目的人物,狭义相对论才开始受到应有的重视。

在处理理论和实验关系上,爱因斯坦为我们提供了光辉的范例。他尊重实验事实,但又不拘泥于个别实验的结果,考夫曼的实验曾一度否定了爱因斯坦的相对质量—速度关系,但是爱因斯坦没有动摇,坚定地相信自己的狭义相对论是“总结了大量现象得到的理论体系”,是经得住考验的。历史证明,真理在他这边。

### 11.2.3 狭义相对论的基本原理

**1. 伽利略变换**

相对于不同的参考系,经典力学定律的形式是完全一样的吗? 牛顿力学的回答:对于任何惯性参考系,牛顿力学的规律都具有相同的形式。这就是经典力学的相对性原理。

经典力学在相对性原理和牛顿的绝对时空基础上,推导出了两个相对作匀速直线运动的坐标关系(参考系)的伽利略变换,即

$$x'=x-ut,\quad y'=y,\quad z=z'\quad t'=t \tag{11-2-1}$$

其中,$u$ 是两个参考系之间的相对速度。伽利略变换是绝对时空概念的直接反应。

由式(11-2-1)可进一步得到速度变换公式

$$v'_x=v_x-u,\quad v'_y=v_y,\quad v'_z=v_z \tag{11-2-2}$$

式(11-2-2)中的三式可以合并成为一个矢量式,即

$$\boldsymbol{v}'=\boldsymbol{v}-\boldsymbol{u} \tag{11-2-3}$$

将式(11-2-3)对时间求导,可得到加速度变换公式,即

$$\boldsymbol{a}'=\boldsymbol{a} \tag{11-2-4}$$

在经典力学中,认为质点的质量与其运动速度无关,即有 $m'=m$;在力学中所见到的质点间的作用力一般都只与质点间的相对位置和相对速度有关,而在伽利略变换下,相对位置和相对速度都是不变量,因而在不同的惯性系中,质点所受到的作用力是相同的,即有 $\boldsymbol{F}'=\boldsymbol{F}$。这样,在惯性系中就有

$$\boldsymbol{F}'=m'\boldsymbol{a}' \tag{11-2-5}$$

这表明牛顿第二定律具有伽利略变换的不变性。

**2. 爱因斯坦的假设**

狭义相对论是建立在“两个基本假设”基础上,彻底改变人类时空观的伟大理论。这两个基本假设的内容如下:

(1) 相对性原理

在所有的惯性系中,一切物理定律都相同,即具有相同的数学表达形式。或者说,对于描述一切物理现象的规律来说,所有惯性系都是等价的。

和牛顿相对性原理加以比较,可以看出前者是后者的推广,使相对性原理不仅适用于力学现象,而且适用于所有物体现象,包括电磁现象在内。这样,我们就可以料到,在任何一个惯性系内,不但是力学实验,而且任何物理实验都不能用来确定本参考系的运动速度。绝对运动或绝对静止的概念,从整个物理学中被排除了。相对性原理将对称性推广于全部基础

物理学。

（2）光速不变原理

在所有惯性系中，真空中光沿各方向传播的速率都等于一个恒量 $c$，与光源和观察者的运动状态无关。

狭义相对论的这两条基本假设虽然非常简单，但是却和人们已经习以为常的经典时空观及经典力学体系不相容。确定两个基本假设就必须彻底地摒弃绝对时空观念，修改伽利略坐标变换的力学定律，爱因斯坦从考虑同时性的相对性开始，得到一套新的时空变换公式——洛伦兹变换。

**3. 洛伦兹变换**

在本书1.5节只是直接给出洛伦兹变换式。下面我们根据爱因斯坦的相对论时空概念，以光速不变原理为依据，推导出相对论坐标变换关系——洛伦兹变换。

为简便起见，如图11-2-1所示，我们假设 $S$ 系和 $S'$ 系是两个相对作匀速直线运动的惯性坐标系。

规定 $S'$ 系沿 $S$ 系的 $x$ 轴正方向以速度 $u$ 相对于 $S$ 系作匀速直线运动，$x'$、$y'$ 和 $z'$ 轴分别与 $x$、$y$ 和 $z$ 轴平行，$S$ 系原点 $O$ 与 $S'$ 系原点 $O'$ 重合时两惯性坐标系在原点处的时钟都指示零点。

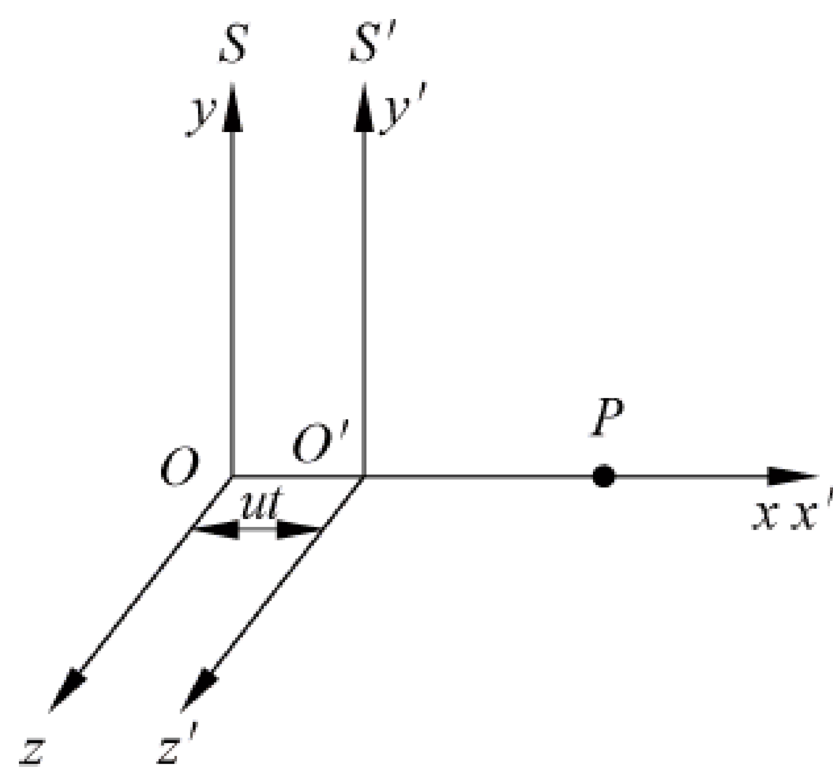



图 11-2-1　时空坐标的变换

新变换首先应该满足狭义相对论的两条基本原理。另外，当运动速度远小于真空光速时，新变换应该过渡到伽利略变换，因为在这种情况下伽利略变换被实践检验是正确的。最后，新变换应该是线性的，因为只有这样才能保证当物体在一个参考系中作匀速直线运动时，在另一个参考系中也观察到它作匀速直线运动。

根据这些要求，我们作最简单的假设

$$x' = k(x - ut) \tag{11-2-6}$$

$k$ 是比例系数，与 $x$ 和 $t$ 都无关。按照狭义相对论的第一条基本原理，$S$ 系和 $S'$ 系除了作相对运动外别无差异，考虑到运动的相对性，相应地，应有

$$x = k(x' + ut') \tag{11-2-7}$$

另外两个坐标的变换容易写出

$$y' = y \tag{11-2-8}$$

$$z' = z \tag{11-2-9}$$

为得到时间坐标的变换，将式(11-2-6)代入式(11-2-7)，得 $x = k^2(x - ut) + kut'$，从中解出 $t'$，得

$$t' = kt + \left(\frac{1 - k^2}{ku}\right)x \tag{11-2-10}$$

确定 $k$ 需要用到狭义相对论的第二条基本原理。根据我们规定的初始条件，当两个惯性坐标系的原点重合时，有 $t = t' = 0$。如果就在这时，在共同的原点处有一点光源发出一光脉冲，在 $S$ 系和 $S'$ 系都观察到光脉冲以速率 $c$ 向各个方向传播。所以在 $S$ 系

$$x = ct \tag{11-2-11}$$

在 $S'$ 系有

$$x' = ct' \tag{11-2-12}$$

将式(11-2-11)和式(11-2-12)代入式(11-2-6)和式(11-2-7),得

$$ct' = k(c-u)t \tag{11-2-13}$$

和

$$ct = k(c+u)t' \tag{11-2-14}$$

由以上两式消去 $t$ 和 $t'$ 后,可解得

$$k = \frac{1}{\sqrt{1-u^2/c^2}} \tag{11-2-15}$$

将 $k$ 代入式(11-2-6)和式(11-2-7)就得到新变换的最终形式

$$\left.\begin{aligned} x' &= \frac{x-ut}{\sqrt{1-u^2/c^2}} \\ y' &= y \\ z' &= z \\ t' &= \frac{t-ut'}{\sqrt{1-u^2/c^2}} \end{aligned}\right\} \tag{11-2-16}$$

$$\left.\begin{aligned} x &= \frac{x'+ut'}{\sqrt{1-u^2/c^2}} \\ y &= y' \\ z &= z' \\ t &= \frac{t'+ux'/c^2}{\sqrt{1-u^2/c^2}} \end{aligned}\right\} \tag{11-2-17}$$

根据相对性原理,将式(11-2-16)中带撇的量与不带撇的量互换,并将 $u$ 换成 $-u$,就得到洛伦兹变换的逆变换式(11-2-17)。这种新的变换称为洛伦兹变换。

**4. 时间延缓与孪生子佯缪**

前面我们只对时间相对性作了定性的讨论,下面我们来讨论时间度量和参考系相对速度之间的关系。

如图 11-2-2 所示,假设在 $S'$ 系 $A'$ 点处内有一闪光光源,紧挨着它有一只标准钟 $C'$,正对面放置一面反光镜 $M'$,可使光源发出的光脉冲原路返回。设反射镜与 $x$ 轴相距为 $d$,则光脉冲从 $A'$ 出发再回到 $A'$ 这两个事件的时间间隔由钟 $C'$ 给出,即

$$\Delta t = \frac{2d}{c} \tag{11-2-18}$$

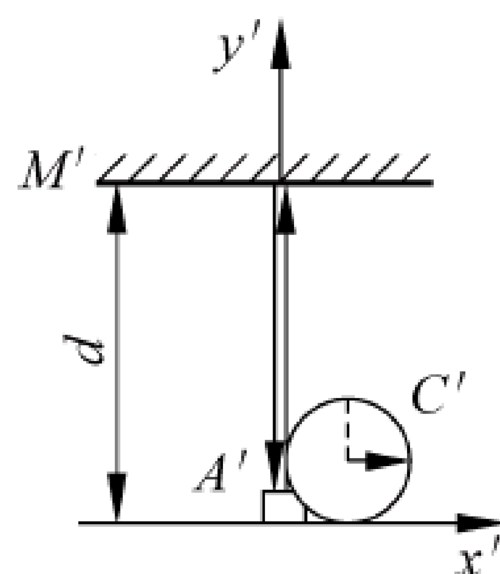


图 11-2-2 在 $S'$ 系中测量时间

如图 11-2-3 所示,从 $S$ 系观察,由于 $S'$ 系的运动,这两个事件不发生在 $S$ 系中的同一地点。为了测量这一时间间隔,必须利用沿 $x$ 轴配置的许多静止于 $S$ 系的经过校准而同步的钟 $C_1, C_2$ 等,而待测的时间间隔由光从 $A'$ 发出和返回 $A'$ 时,$A'$ 所邻近的钟 $C_1, C_2$ 给出。在 $S$ 系中,由于光源在运动,光线走的是锯齿形路径(图 11-2-3(b))。光线"来回"一次的时间为

$$\Delta t = \frac{2l}{c} = \frac{2}{c}\sqrt{d^2 + \left(\frac{u\Delta t}{2}\right)^2} \tag{11-2-19}$$

由此式解出 $\Delta t=\frac{2l}{c}\frac{1}{\sqrt{1-u^2/c^2}}$，和式(11-2-18)比较可得

$$\Delta t=\frac{\Delta t'}{\sqrt{1-u^2/c^2}} \tag{11-2-20}$$

任何物体运动的速度都不可能超过光速，所以由式(11-2-20)可知 $\Delta t>\Delta t'$。即在一个惯性系(如上述 $S$ 系)中，运动的钟比静止的钟走得更慢。这种效应叫作运动时钟延缓或慢钟效应。

必须指出的是，在式(11-2-19)中，我们用到了在参考系 $S$ 中镜面到 $x$ 轴的距离仍然是 $d$，这可由下述假想实验得出。

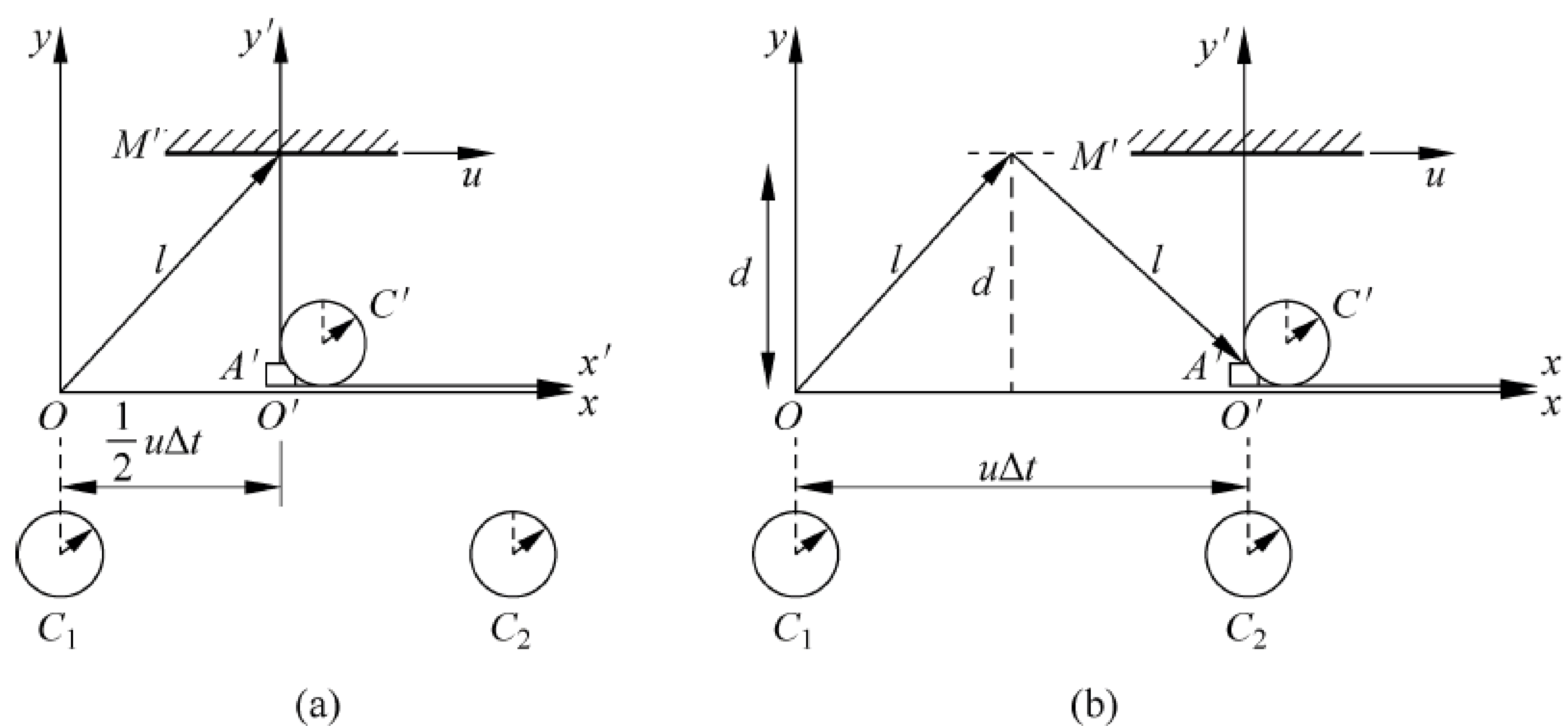



图 11-2-3 在 $S$ 系中测量：时间度量与参考系相对速度的关系

设山洞外停有一列火车，列车高度与山洞高度相等。现在使列车匀速向山洞开去。这时列车的高度是否和洞顶高度相等呢？假设列车高度由于运动而变小了，那么在地面上观察，由于运动列车高度变小，列车当然能顺利通过山洞。如果在列车上观察，则山洞是运动的，由相对性原理，洞顶的高度应减小，这样列车不能通过山洞。这就发生了矛盾。但是列车能否通过山洞是一个确定的物理事实，和参考系的选择无关，即上述矛盾不应该发生。这说明上述假设是错误的。因此满足相对性原理的条件下，列车和山洞的高度不因运动而减小，即垂直于相对运动方向的长度测量与运动无关。

此外还应注意，运动是相对的。$S$ 系里的人看 $S'$系里的钟变慢，而 $S'$系里的人看 $S$ 系里的钟也比自己慢。我们将相对于物体(或观察者)静止的钟所显示的时间间隔 $\Delta\tau$ 叫作该物体的固有时。式(11-2-20)中 $\Delta t'$就是光从 $A'$出发又返回 $A'$所经历的固有时 $\Delta\tau$。由式(11-2-20)可以看出，固有时最短。由式(11-2-20)还可以看出，当 $u\ll c$ 时，$\sqrt{1-u^2/c^2}\approx 1$，则 $\Delta t\approx\Delta t'$。这种情况下，两个事件之间的时间间隔在各参考系中测得的结果都是一样的，即时间的测量与参考系无关。这就是牛顿的绝对时间的概念。由此可知，牛顿的绝对时间概念实际上是相对论时间概念在参考系的相对速度很小时的近似。

在相对论中，每位观察者都具有自身的时间测度。这会导致所谓的孪生子佯缪(twin paradox)。孪生子中的哥哥(a)出发进行空间航行，其航行的速度接近光速，而孪生子中的弟弟(b)留在地球上。根据相对论效应，高速运动的时钟变慢，等哥哥返回与弟弟相遇时，弟弟已经变得很老了，因为地球上已经过了几十年了。这真应了古代神话里“天上方一日，地上已千年”的说法。然而按照相对性原理，飞船相对于地球作高速运动，反过来地球相对

于飞船也作高速运动,弟弟看哥哥年轻了,哥哥看弟弟也应该年轻了。也就是说,若假定地球为静止参考系,由相对论时间膨胀公式可得:$t_a=\dfrac{t_b}{\sqrt{1-u^2/c^2}}$;但是我们也可以假定火箭是"静止"的参考系,那么则得到另一种时间关系式:$t_b=\dfrac{t_a}{\sqrt{1-u^2/c^2}}$。

为什么在这里天(飞船)、地(地球)两个参考系不对称?

从逻辑上看,这佯谬并不存在。首先,他们两者的结果不会都是对的,即两个人再相遇时,不会都说对方比自己年轻,因为物理上的现实结果只有一个。其次,因为飞船、地球两个参考系的确是不对称的。以近光速飞行的哥哥在离开地球和返回地球的途中经历了加速和减速过程,这样就不能认为哥哥经历的运动与弟弟的运动相互对称了。由于飞船的加速和减速过程,因而与他相联系的参考系就不可能总是惯性系,在同他相联系的参考系中进行讨论就必定要用到加速参考系中的规律,这超出了狭义相对论的理论范围,需要用广义相对论去讨论。所以,以地球这个惯性系为参考系的时钟延缓效应的计算是正确的,而以飞船这个非惯性系为参考系的计算是错误的,即旅行中的哥哥确实更年轻些。

实际上,通过原子钟环球航行,运动中的放射性粒子平均衰变寿命的测定等实验,可以实际观测到这个效应。原子钟环绕地球飞行时,向东飞行的原子钟发现地面上的原子钟走得慢一些,而向西飞行的原子钟则比地面上的钟要快一些。这是因为向东飞行的原子钟的实际运动速度,大于向西飞行的原子钟的运动速度。

**5. 长度收缩**

上面我们谈到了时间的相对性问题,其实,光速不变原理还会带来空间长度的相对性问题。那就是说,同一物体的长度,在不同的参考系内测量,会得到不同的结果。在前面我们已经说明了垂直于运动方向的长度测量与参考系无关,那么沿运动方向的长度应又如何呢?

通常,在某个参考系中,一个静止的物体的长度可以由一个静止的观察者用尺去量。但要测量一个运动物体的长度,就必须在某一时刻同时(本参考系中的同时)确定物体两端在参考系中的位置。如图 11-2-4 所示要测量一个运动小车的长度。合理的办法是:在同一时刻,同时记下小车车头位置 $x_1$ 和车尾位置 $x_2$,那么运动小车的长度 $l=x_2-x_1$(图 11-2-4(a))。如果 $x_1$,$x_2$ 这两个位置不是同时记录,而是在记录 $x_1$ 后过了一段时间再记录 $x_2$(图 11-2-4(b)),则 $x_2-x_1$ 就不能代表小车的长度,而且 $x_2-x_1$ 的值和两次记录的时间间隔有关系。

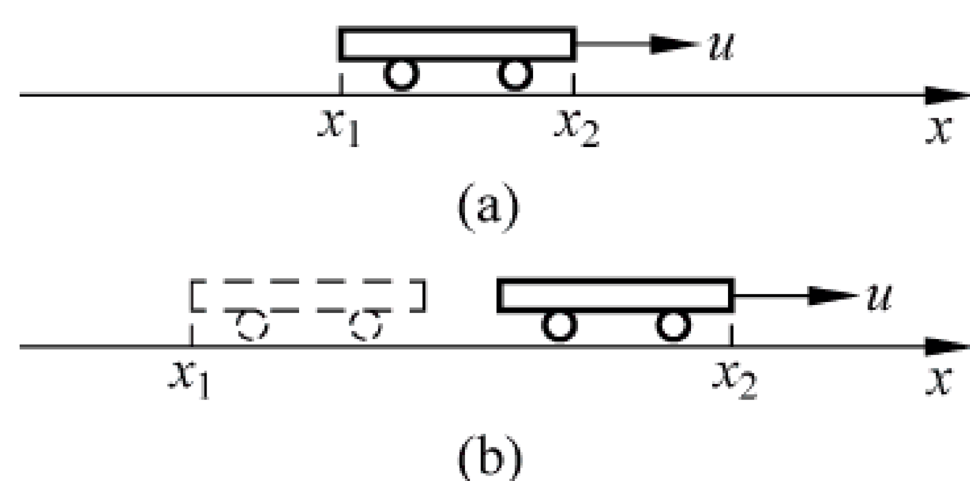



图 11-2-4 测量运动小车的长度

(a) 同时记录 $x_1$,$x_2$;(b) 同时记录 $x_1$,$x_2$

可见,测量一个运动物体的长度是和同时性概念密切相关的,也就是和时间的测量密切联系起来了。

如果一辆小车静止在 $S'$系中,小车的车尾处有一个光源,车头处放置一面反光镜。在 $S'$系中看小车的长度为 $L$。若从车尾向车头发射一束光,则这束光返回所用的时间为 $t'=2L'/c$;从 $S$ 系看,由于小车运动速度为 $u$,故光从车尾传到车头的时间为 $t_1=L/(c-u)$;光从车头返回车尾的时间为 $t_2=L/(c+u)$。总时间是 $t=t_1+t_2=2L/c(1-u^2/c^2)$。由于 $t'$是固有时,所以有 $t=\dfrac{t'}{\sqrt{1-u^2/c^2}}$,即 $t=\dfrac{2L}{c(1-u^2/c^2)}=\dfrac{2L'}{c\sqrt{1-u^2/c^2}}$,所以得

$$L=L'\sqrt{1-u^2/c^2} \tag{11-2-21}$$

我们注意到，在参考系 $S'$（小车参考系）中测得的小车的长度 $L'$ 小于在地面参考系中测得长度 $L$，即运动小车（纵向）的长度收缩，也称为洛伦兹收缩，且长度收缩只发生于沿运动的方向。我们把与物体相对静止的参考系中测出的长度叫物体的固有长度，以区别于它运动时的长度。那么式(11-2-21)中的 $L'$ 就是固有长度。

**6. 相对论速度变换**

设同一质点在 $S$ 系和 $S'$ 系中的速度分别为 $v$ 和 $v'$。

$$S: v_x=\frac{\mathrm{d}x}{\mathrm{d}t},\quad v_y=\frac{\mathrm{d}y}{\mathrm{d}t},\quad v_z=\frac{\mathrm{d}z}{\mathrm{d}t};\qquad S': v'_x=\frac{\mathrm{d}x'}{\mathrm{d}t'},\quad v'_y=\frac{\mathrm{d}y'}{\mathrm{d}t'},\quad v'_z=\frac{\mathrm{d}z'}{\mathrm{d}t'}$$

取洛伦兹变换的微分，可得相对论速度变换关系

$$\left.\begin{aligned} v'_x&=\frac{v_x-u}{1-uv_x/c^2}\\ v'_y&=\frac{v_y\sqrt{1-u^2/c^2}}{1-uv_x/c^2}\\ v'_z&=\frac{v_z\sqrt{1-u^2/c^2}}{1-uv_x/c^2}\end{aligned}\right\} \tag{11-2-22}$$

正变换

$$\left.\begin{aligned} v_x&=\frac{v'_x+u}{1+uv'_x/c^2}\\ v_y&=\frac{v'_y\sqrt{1-u^2/c^2}}{1+uv'_x/c^2}\\ v_z&=\frac{v'_z\sqrt{1-u^2/c^2}}{1+uv'_x/c^2}\end{aligned}\right\} \tag{11-2-23}$$

逆变换

从中我们看到，虽然垂直于运动方向的长度不变，但速度是变的，这是因为时间间隔变了。易见，$u\ll c$，$v_x\ll c$，式(11-2-22)简化为

$$v'_x=v_x-u,\quad v'_y=v_y,\quad v'_z=v_z \tag{11-2-24}$$

这就是我们熟知的经典速度合成公式，具体推导过程请读者参考本书1.5节。

**7. 狭义相对论的动力学**

在动力学中有一系列物理概念，如动量、质量、速度、能量、力、功等。这些物理量在相对论中都面临着重新定义的问题。下面就介绍相对论动力学的基本内容。

1）相对论质量

在相对论中，一个质点的动量 $p$ 仍定义为

$$\boldsymbol{p}=m\boldsymbol{v} \tag{11-2-25}$$

我们把上式中动量与速度的比例系数 $m$ 仍定义为该质点的质量。可以证明，要使动量守恒定律在洛伦兹变换下保持形式不变，则质点的质量 $m$ 不再认为是一个与其速率无关的常量，而是随速率增大而增大，且应认为

$$m=m(v)=\frac{m_0}{\sqrt{1-u^2/c^2}} \tag{11-2-26}$$

式中 $m_0$ 是质点静止时的质量，即由相对该质点静止的观察者测得的质量，称为静质量。此式表明，当质点以一定速率相对观察者运动，即由相对该质点运动的观察者测量时，其质量 $m$ 大于静止质量 $m_0$。因此，质点的质量也是相对的。式(11-2-26)称为相对论的质速关系式。

图 11-2-5 示出了$\dfrac{m}{m_0}\sim\dfrac{u}{c}$关系曲线，可以看出，当质点速度接近光速时，其质量变得很大，欲使之再加速就很困难，这就是一切物体的速率都不可能达到和超过光速 $c$ 的动力学原因。实验证明，在高能加速器中的粒子，随着能量的大幅度增加，其速率只是越来越接近光速，而从来没有达到或超过真空中的光速 $c$。实验还证实了式(11-2-26)的正确性。

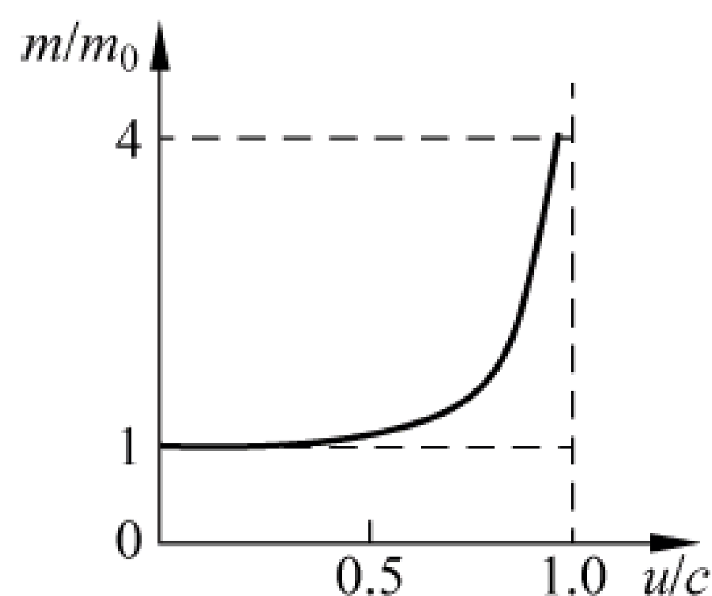


图 11-2-5 质量随速率变化

由式(11-2-26)可得到质点的相对论质量 $\boldsymbol{p}$ 与其速度 $\boldsymbol{v}$ 的关系为

$$\boldsymbol{p}=m\boldsymbol{v}=\frac{m_0}{\sqrt{1-u^2/c^2}}\boldsymbol{v} \tag{11-2-27}$$

还可以证明，对洛伦兹变换保持形式不变的相对论质点动力学方程为

$$\boldsymbol{F}=\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{m_0}{\sqrt{1-u^2/c^2}}\boldsymbol{v}\right] \tag{11-2-28}$$

(1) $v\ll c$ 时，$\boldsymbol{F}=\dfrac{\mathrm{d}}{\mathrm{d}t}(m_0\boldsymbol{v})=m_0\dfrac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=m_0a\rightarrow$牛顿第二定律。

(2) $\boldsymbol{F}=\dfrac{\mathrm{d}}{\mathrm{d}t}(m\boldsymbol{v})=m\dfrac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}+\dfrac{\mathrm{d}m}{\mathrm{d}t}\boldsymbol{v}$，$\boldsymbol{F}$ 不仅改变速度，而且改变质量。

(3) 当 $v\rightarrow c$ 时，$\dfrac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}\rightarrow 0$，不再改变，所以光速 $c$ 是极限速度。

(4) 由 $m=\dfrac{m_0}{\sqrt{1-u^2/c^2}}$知，当 $v\rightarrow c$ 时，所以必须有 $m_0=0$。即光子的质量为零，没有静止的光子。

2) 相对论动能

在经典力学中，质点动能表示式为 $E_\mathrm{k}=\dfrac{1}{2}mv^2$，式中 $m$ 为常数，并且质点动能的增加等于合力对质点所做的功。在相对论的力学中，我们认为动能定理仍然适用，并且由此可以导出相对论中质点动能的表达式。若取速度为零时，质点动能为零，则在力 $\boldsymbol{F}$ 作用下，质点速率由零增大到 $v$ 时，其动能为

$$E_\mathrm{k}=\int\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}=\int_0^v\mathrm{d}(m\boldsymbol{v})\cdot\boldsymbol{v} \tag{11-2-29}$$

其中，$\mathrm{d}(m\boldsymbol{v})\cdot\boldsymbol{v}=\mathrm{d}m\boldsymbol{v}\cdot\boldsymbol{v}+m\mathrm{d}\boldsymbol{v}\cdot\boldsymbol{v}=v^2\mathrm{d}m+mv\mathrm{d}v$，又由质速关系式(11-2-25)可得：$m^2v^2=m^2c^2-m_0^2c^2$，对等式两边取微分，并整理即得：$v^2\mathrm{d}m+mv\mathrm{d}v=c^2\mathrm{d}m$。将此结果代入式(11-2-29)，即得质点的相对论动能作为速率函数的表达式

$$E_\mathrm{k}=\int_{m_0}^{m}c^2\mathrm{d}m=mc^2-m_0c^2 \tag{11-2-30}$$

初看起来，它和经典的质点动能表达式全然不同，但当 $v\ll c$ 时，有 $\sqrt{1-u^2/c^2}\approx$

$1+\frac{1}{2}\cdot\frac{u^2}{c^2}$；代入式(11-2-30)，即得质点作低速运动时的动能为 $E_k=\frac{1}{2}m_0v^2$。这就和经典动能的表达式相同了。

3）相对论能量及质能关系

(1) 相对论能量

在相对论动能公式 $E_k=mc^2-m_0c^2$ 中，由于等号右端两项都具有能量的量纲，故将式(11-2-30)写成

$$mc^2-m_0c^2=E_{总}-E_0 \tag{11-2-31}$$

其中，认为 $m_0c^2$ 表示质点静止时具有的能量，叫静能。而 $mc^2$ 表示质点以速率 $v$ 运动时所具有的能量，这个能量是在相对论意义上质点的总能量，二者之差即为质点由于其运动而增加了的能量，也就是动能 $E_k$。式(11-2-30)表明质点的动能等于质点该时刻的总能量和静能之差。

(2) 相对论质量与能量关系的讨论

按照相对论的概念，质点在相互作用(如碰撞)过程中，最一般的能量守恒应表示为

$$\sum_i E_i=\sum_i(m_ic^2)=常量 \tag{11-2-32}$$

由此公式立即可以得出，在相互作用过程中

$$\sum_i m_i=常量 \tag{11-2-33}$$

这表示质量守恒。在相对论中，将历史上分别发现的两条相互独立的自然规律完全统一起来了。相对论能量和质量守恒是一个统一的物理规律。

质点质量和能量之间的关系由质能关系式(11-2-34)决定。

$$\left.\begin{aligned}E&=mc^2\\E_0&=m_0c^2\end{aligned}\right\} \tag{11-2-34}$$

质能关系式揭示出质量和能量这两个物质基本属性之间的内在联系，即一定质量 $m$ 相应地联系着一定的能量 $E=mc^2$，即使处于静止状态的物体也具有能量 $E_0=m_0c^2$。它的提出具有划时代的意义，它开创了原子能时代。

由 $E=mc^2$ 表示，质量和能量的不可分割性。

$$\Delta E=\Delta(mc^2)=c^2\Delta m \tag{11-2-35}$$

由式(11-2-35)可知，物体质量发生 $\Delta m$ 变化，物体的能量也一定有相应的变化。

当物体静止时，仍然存在静能：$E_0=m_0c^2$，所以物体的静能是物体内能的总和：分子运动动能、分子间相互作用势能、分子内部各原子动能和相互作用势能、原子内部使核和电子结合在一起形成原子的电磁能等。

因为 $c$ 很大，所以即使 $m_0$ 很小，静止能量 $E_0=m_0c^2$ 也会很大，即质能关系式预言了物质的质量就是能量的一种储藏。

在某些原子核反应中，如重核裂变和轻核聚变过程中，会发生静止质量减小的现象，称为质量亏损。由质能关系式可知，这时静止能量也相应地减少。但在任何过程中，总质量和总能量又是守恒的，因此这意味着，有一部分静止能量转化为反应后粒子所具有的动能。而后者又可以通过适当方式转变为其他形式能量释放出来，这就是某些裂变和核聚变反应能够释放出巨大能量的原因。

质能关系式为人类利用核能奠定了理论基础，它是狭义相对论对人类最重要的贡献之一。

4）相对论能量和动量的关系

为了找到能量和动量之间的关系，我们将式(11-2-26)两边平方乘 $c^2$：$m^2c^2\left(1-\frac{v^2}{c^2}\right)=m_0^2c^2$ 即 $m^2c^2(c^2-v^2)=m_0^2c^4$。由于 $\boldsymbol{p}=m\boldsymbol{v}$ 上式又可以写成

$$E^2=p^2c^2+m_0^2c^4 \tag{11-2-36}$$

这就是相对论中同一质点的能量和动量之间的关系式。以 $E$、$pc$、$m_0c^2$ 表示三角形的三边。

(1) $v\ll c\Rightarrow E_k\ll m_0c^2$ 略去 $E_k^2$，得到 $E_k=p^2/2m_0\to$ 回到了牛顿力学。

(2) 对于光子，则有 $E=pc$，于是可知光子的动量为

$$p=\frac{h\nu}{c}=\frac{h}{\lambda} \tag{11-2-37}$$

### 11.2.4 狭义相对论的意义

狭义相对论变革了传统的时间和空间的概念；在狭义相对性原理的基础上统一了牛顿力学和麦克斯韦电动力学两个体系，指出它们都服从狭义相对性原理，都是对洛伦兹变换协变的，给出了物体在高速运动下的运动规律，牛顿力学只不过是物体在低速运动下很好的近似规律。

狭义相对论揭示了质量和能量的统一，提出了质量与能量相当，给出了质能关系式。这两项成果对低速运动的宏观物体并不明显，但在研究微观粒子时却显示了极端的重要性。因为微观粒子的运动速度一般都比较快，有的接近甚至达到光速，所以粒子的物理学离不开相对论。质能关系式不仅为量子理论的建立和发展创造了必要的条件，而且为原子核物理学的发展和应用提供了根据。

狭义相对论与10年之后创立的广义相对论，在科学史上矗立了一座巍峨而永恒的丰碑。相对论是现代物理学的重要基石。它的建立是20世纪自然科学最伟大的发现之一，对物理学、天文学乃至哲学思想都有深远影响。

相对论是科学技术发展到一定阶段的必然产物，是电磁理论合乎逻辑的继续和发展，是物理学各有关分支又一次综合的结果。相对论从逻辑思想上统一了经典物理学，使经典物理学成为一个完美的科学体系。

自相对论建立以来，已经过去了很长时间，它经受住了实践和历史的考验，是人们普遍承认的理论。

## 11.3 广义相对论

从前面的讨论中我们知道，在狭义相对论诞生之前，有许多科学家进行了大量的探索性研究，虽然史料表明这些研究对爱因斯坦建立狭义相对论似乎没有太直接和重要的影响，那么广义相对论物理思想可以说是爱因斯坦一人的成就。

广义相对论是爱因斯坦于1915年以几何语言建立的引力理论，综合了狭义相对论和牛顿的万有引力定律，将引力描述成因时空中的物质与能量而弯曲的时空，以取代传统对于引力是一种力的看法。因此，狭义相对论和万有引力定律，都只是广义相对论在特殊情况之下

的特例。

### 11.3.1 发展历史

**1. 狭义相对论的局限性**

爱因斯坦于1905年9月在《物理纪事》上发表了《论动体的电动力学》这篇划时代的论文，宣告了狭义相对论的诞生。但是狭义相对论建立以后，爱因斯坦并没有止步。他认为狭义相对论还有许多问题没有解决。还存在着以下的局限性。

(1) 狭义相对论不能解释惯性系的特殊、优越性问题

在经典力学中，力学方程在不同的惯性参考系中具有不变性，若将同一物理现象放在非惯性系中描述，其力学方程将发生改变，这就产生了一个问题："惯性参考系是否在物理上具有特殊的性质和较其他参考系更为优越呢?"牛顿和马赫(Ernst Mach，1838—1916)都曾研究过这个问题，但都无法说明同一物理现象对惯性系和非惯性系为什么会有不同的表现。

为什么惯性坐标系在物理学中比其他坐标系更为优越？为什么一切物体在引力场中下落都具有同样的加速度？在狭义相对论里，为什么惯性质量随能量变化？刚刚经受住考验的狭义相对论，为什么一用到引力场中就遇到了矛盾？爱因斯坦感到极大的疑惑。爱因斯坦发现在狭义相对论的范畴内，仍然无法解决惯性系的特殊、优越性问题。

爱因斯坦从牛顿、马赫的失败中认识到，从理论上论证惯性系的特殊和优越性是不可取的，狭义相对论原理对这个理论问题同样做不出解答。爱因斯坦尖锐地意识到这一点，并认为这是经典力学、狭义相对论都"固有的认识论上的缺点"。

爱因斯坦坚信自然界的和谐与统一，认为对惯性坐标系的优越做出解释，出路只有两条：要么从理论上阐明惯性坐标系特殊、优越的物理原因；要么从理论上消除惯性坐标系特殊、优越的物理地位。爱因斯坦选择了第二条出路，就是扩大狭义相对论原理的物理内容。1933年爱因斯坦曾回忆说："如果速度概念只能有相对的意义，难道我们还固执地把加速度当作一个绝对的概念吗?"他思索着从扩大狭义相对论的内容和原理着手，用新的观念和论点从理论上消除惯性系的特殊和优越。

(2)"刚性"空-时度规不能解决非惯性系中的问题

在经典力学中，时空是绝对的，几何学只是一种数学工具，反映了一个平直的物理空间。在狭义相对论中，光速是恒定的，以光速为基础的时空也是均匀的、各向同性的，狭义相对论所依附的四维时空统一体是一个具有平直性质欧几里得空间。

狭义相对论虽然摆脱了绝对时空观的束缚，确立了同时性的相对性，时间间隔的相对性和空间长度测量与参考系运动状态有关等新的时空观。然而，在同一参考系中，仍按统一的测量标准测量时间与长度，这意味着狭义相对论中仍具有刚性的尺和同步的时钟，空-时度规的性质仍是"刚性"的。

爱因斯坦在研究中发现，当我们处在一个非惯性系内的时候，大范畴内平直的坐标空间概念已经失效，不能够用到处同步的钟和到处一样长度且与取向无关的尺，狭义相对论所依附的空-时度规不能解决非惯性系的问题。

(3) 狭义相对论不能解决引力现象问题

爱因斯坦曾反复运用数学方法修改牛顿的引力理论，企图把引力现象归纳在狭义相对论的范畴之内，但没有获得成功。爱因斯坦通过狭义相对论，把电场与磁场统一起来，把质量和能量统一起来，把牛顿力学方程作了相对论修正，使之与麦克斯韦方程协调起来了。接着，爱因斯坦想把引力现象也纳入狭义相对论的范畴。他发现：物体的质量是它所含能量

的量度,若物体的能量改变 $\varepsilon$,那么质量也要改变 $\varepsilon/c^2$。按这个结论,物体的惯性质量将随其能量而变,落体的加速度同它的水平速度或者物体的内能有关。这不符合在引力场中的一切物体都具有同一加速度的实验事实。推出的结论不能解释引力场中的普遍事实。

爱因斯坦在 1907 年开始认识到在狭义相对论的框架里,是不可能有令人满意的引力理论的。

**2. 广义相对论的创建思路分析**

爱因斯坦 1922 年回忆他创建广义相对论的过程时讲道:当他正在思考如何突破狭义相对论的框架,以解决惯性与重量之间的不协调时,一个突然的闪念出现了。他说:“有一天,突破口突然找到了。当时我正坐在伯尔尼专利局办公室里,脑子里突然闪现了一个念头:如果一个人正在自由下落,他决不会感到他有重量。我吃了一惊,这个简单的思想实验给我的印象太深了。它把我引向了引力理论。我继续想下去:下落的人正在作加速运动,可是在这个加速参考系中,他有什么感觉?他如何判断面前所发生的事情?”

爱因斯坦 1933 年在《广义相对论的来源》一文中,这样写道:“在引力场中一切物体都具有同一加速度,这条定律也可以表述为惯性质量与引力质量相等的定律,它当时就使我认识到它的全部重要性。我为它的存在感到极为惊奇,并猜想其中必定有一把可以更加深入地了解惯性和引力的钥匙。”

(1)“等效原理”的建立

对于惯性质量和引力质量,通过伽利略、牛顿、贝塞尔(Bessel Friedrich Wihelm,1784—1864)等所进行的实验研究,使人们得出结论:惯性质量恒等于引力质量。

真空中在一给定地点作自由下落的所有物体加速度相同的事实,表明各种物体的引力质量与惯性质量有严格的正比关系。爱因斯坦对升降机进行了一系列的理想实验之后,基于引力质量等于惯性质量的实验事实,把实验结果归纳为:引力场同参考系相当的加速度在物理上完全等价。从而建立了把相对原理推广到加速度的非惯性系的“等效原理”。

爱因斯坦决心以惯性质量与引力质量相等这一事实出发,寻求新的引力理论。在进一步认识了惯性力的本质是非惯性系加速度的反映之后,爱因斯坦提出了等效原理的假设,他说:“这个假说的启发性意义在于,它允许用一个均匀加速参考系来代替一个均匀引力场,而均匀加速参考系这种情况,从理论研究的观点看来,在一定程度上是可以接受的。”于是,爱因斯坦又把狭义相对性原理作了相应的推广,第一次正式提出:“迄今为止,我们只把相对性原理,即认为自然规律同参考系的状况无关这一假设应用于非加速参考系。是否可以设想,相对性运动原理对于相互作加速运动的参考系也仍然成立?”

他利用均匀加速参考系与均匀引力场的等效性,研究了引力场对时空的影响。1911 年,在《关于引力对光传播的影响》的论文中,应用等效原理,推出了光线在经过太阳附近时要发生弯曲的结论。

(2)“广义协变原理”的建立

在参考系变换下物理规律方程形式不变的性质称为协变性。狭义相对论对于绝对时空观给予了有力的冲击,克服了同时性的绝对性及时间间隔与空间长度的测量与参考系运动状态无关的传统观念。但是,在同一参考系中,仍保持着统一的时间和长度的测量标准,即具有刚性的尺和同步的钟。空间长度、时间间隔的测量,通常与坐标差对应。然而,根据等效原理及加速度运动的相对论,对于物理定律来说,不能仅限于满足洛伦兹变换的协变性——这仅仅对应于速度的相对性;而应扩大为更普遍的非线性变换协变性——这对应着

加速度的相对性以及与引力场之间的等效性，即应把洛伦兹变换推广为非线性变换。爱因斯坦在谈到当时的认识情况时曾写道："接受了等效原理所要求的非线性变换，对于坐标的简单的物理解释无可避免是致命的——那就是说，不能再要求：坐标差应当表示那些用理想标尺或理想时钟所得的直接量度结果。我被这一点知识大大地困惑住了……""这个问题的关键，在于建立柔性的时空度规概念。在这方面，闵可夫斯基关于狭义相对论形成基础的分析显得很重要。"这是从狭义相对论的数学原理表示走向广义相对论数学表示的一个桥梁。

应当指出，实际的引力场并不是均匀的，当将等效原理加以推广时，人们只能将局部点的引力场等效于不变的加速度，而这个加速度的大小从一点到另一点却是变化的。"爱因斯坦升降机"不能在大范围中取代实际的引力场的作用。

要克服这种"刚性"空-时度规的缺陷，爱因斯坦在困难和困惑中转向了数学，研究罗巴切夫斯基（Lobachevsky，1792—1856）、闵可夫斯基、高斯（Gauss，1777—1855）、黎曼（Riemannn，1826—1866）等人的非欧几何。

爱因斯坦在研究中发现，几何学与引力场有密切的内在联系。空-时的属性同物质的分布和运动有关，当没有物质存在时（也没有引力场），空间是平直的（即数学上的欧几里得空间）；当有物质存在时，空间的性质发生了变化，时间和空间变得不均匀，大质量物体的周围空间要发生弯曲（即数学上的黎曼空间），爱因斯坦逐一找到了引力场中的物理量与黎曼几何中的数学量的对应关系。

爱因斯坦在研究中又发现，空-时的度规对引力场有依赖关系。"在过程发生的地点引力势越大，在时钟里所发生的过程（任何物理过程）也就进行得愈快"；"量杆的长度与放置的位置和方向有关"；"光线被引力场所弯曲"；"所有的电磁现象均受引力场的影响"。得出了任何引力场的出现，意味着空-时度规的改变，引力场以完全确定的方式给测量工具和时钟以影响的结论。从而将狭义相对论所依附的"刚性"空-时度规推广到非惯性系和引力场相联系的"柔性"空-时度规（即在非欧空间中随位置改变的黎曼度规）。

爱因斯坦运用"柔性"空-时度规按等效原理的要求把相对性原理推广到非惯性系，在理论上表述为"广义协变原理"："自然界的普遍规律，是由那些对一切坐标系都有效的——即对于无论哪种坐标代换都是协变的方程式来表示的。"现在一般记述为：自然定律在任何参考系中都可以表示为相同的数学形式。

爱因斯坦把惯性系的"刚性"时空推广到加速参考系或引力场的"柔性"时空，完成了对时空认识上新的飞跃。

（3）引力场方程建立的思考

在理论有了突破以后，为了推广相对论，1912年爱因斯坦邀请他的老同学苏黎世工业大学数学教授马尔塞耳·格罗斯曼投入了研究。

1912年，爱因斯坦已建立起"柔性"度规的概念，并明确该种"柔性"度规依赖于引力场。接下来的任务是寻求对于时空"柔性"度规来说满足广义协变性要求的微分方程。在格罗斯曼（Grossman，1905—　）的合作下，爱因斯坦把绝对微积分，即黎曼张量运算引入了物理学，把平直空间的张量运算推广到弯曲的黎曼空间，为建立广义协变性引力理论开辟了道路。

经过近一年的努力，1913年，爱因斯坦与格罗斯曼发表了著名论文《广义相对论与引力纲要》，在这篇论文中，爱因斯坦在肯定时空度规依赖于引力场的思想基础上，运用推广的方

法，给出了一个引力方程。在运用推广方法的过程中，他强调了两点原则：对应原则和逻辑简单性原则。但是爱因斯坦对自己的研究成果并不满意，他以引力势中牛顿-泊松方法为出发点，经过不断地探索、反复地修正，经过一年多极其复杂的计算，终于在1915年建立了重力场方程，1917年提出了具有宇宙因子项重力场方程的普遍形式。

同年3月25日爱因斯坦在普鲁士科学院作了名为《引力的场方程》的学术报告。11月发表了重要论文《用广义相对论解释水星近日点运动》，文中把水星绕日运动当作在太阳引力场中(弯曲室间内)的短程线运动，从定性和定量角度解释了近日点相对于空间某方位的进动。

**3. 广义相对论的建立和三大验证**

1915年爱因斯坦最后建成广义相对论。1916年爱因斯坦发表了论文《广义相对论的基础》，对广义相对论的研究作了全面总结，提出了两个基本原理：等效原理和广义协变原理。

爱因斯坦创造性的思想体现在广义相对论中：物质的存在改变了物理时空的平直性质；空间、时间是弯曲的；时空的弯曲程度反映了引力作用的强弱。广义相对论终于把时间、空间和物体统一在一起了。

广义相对论的引力论在本质上完全不同于牛顿的引力论，人们根据日常的生活经验却分辨不出它们的差异。为了判别它们的差别，爱因斯坦于1915年11月提出三个可供实验验证的推论。即广义相对论三大经典性的检验：

(1) 水星近日点的进动。这一结论已在1859年被法国天文学家勒未里埃(U. J. J. Leverrier，1811—1877)发现。

(2) 光谱线向红端移动的引力红移效应。在强引力场中，时钟要走得慢些，因而从巨大质量的星体表面射到地球上的光的谱线，将显得要向光谱的红端移动。这一结论最终被美国人庞德(R. V. Pound)等人精心组织的实验所证实。

(3) 光线在引力场中偏转(弯曲)。如果光线经过太阳边缘，光线就要发生偏转。爱因斯坦希望天文学家在日全食时进行观测，以验证这一预言。这一理论观点得到了英国天文学家爱丁顿(A. S. Eddingto)的高度重视。他在1919年率领观测队到西非几内亚湾的普林西比岛观测当年5月29日的日全食，同时英国皇家天文学会派了另一支观测队到巴西的索布腊尔。两地的观测结果于1919年11月6日公布，证实了爱因斯坦的预言。全世界为之轰动。这一事件，对于第一次世界大战后恢复国际文化交流起了很大的推动作用。

广义相对论的三大经典性的检验分别被证实与广义相对论的理论推论相符。另外，20世纪60年代的雷达信号经过太阳附近时的延迟和70年代双星因引力波辐射损失能量而使双星运动周期稳定变短的观察，都证实了爱因斯坦理论的正确性。

### 11.3.2 基本原理、理论结构

**1. 基本原理的讨论**

(1) 等效原理——关于质量的两种不同表述

通俗地讲，我们可以用两种方法度量质量，要么我们称它的重量(非常简单)，要么我们测量它对加速度的抵抗(使用牛顿定律)。

人们做了许多实验以测量同一物体的惯性质量和引力质量。所有的实验结果都得出同一结论：惯性质量等于引力质量。

牛顿自己意识到这种质量的等同性是由某种他的理论不能够解释的原因引起的，但他认为这一结果是一种简单的巧合。与此相反，爱因斯坦发现这种等同性中存在着一条取代

牛顿理论的通道。

日常经验验证了这一等同性：两个物体(一轻一重)会以相同的速度“下落”。然而重的物体受到的地球引力比轻的物体大。那么为什么它不会“落”得更快呢？因为它对加速度的抵抗更强。结论是，引力场中物体的加速度与其质量无关。伽利略是第一个注意到此现象的人。重要的是你应该明白，引力场中所有的物体“以同一速度下落”是(经典力学中)惯性质量和引力质量等同的结果。

(2) 广义协变原理——爱因斯坦电梯理想实验

爱因斯坦的广义协变原理可以这样表述：自然法则在所有的系中都是相同的。

电梯理想实验。让我们假想一个在摩天大楼内部自由下落的电梯，里面有一个人(图 11-3-1)。这人让他的表和手绢同时落下。对于一个电梯外以地球为参考系的人来说，表、手绢、人和电梯正以完全一致的速度下落。依据等同性原理，引力场中物体的运动不依赖于它的质量。所以表和地板、手绢和地板、人和表以及人和手绢的距离固定不变。因此对于电梯里的人而言，表和手绢将待在他刚才扔它们的地方。

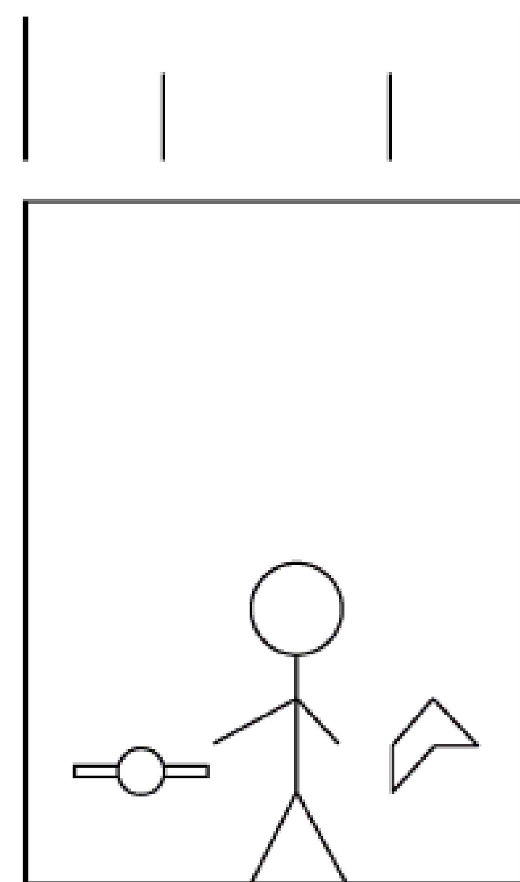


图 11-3-1 加速场与引力场等效

如果这人给他的手表或他的手绢一个特定的速度，它们将以恒定的速度沿直线运动。电梯表现得像一个伽利略系。

现在我们来做第二个理想化的实验：我们的电梯远离任何大质量的物体。比如，正在宇宙深处。突然一个生物开始拖动这个电梯。经典力学告诉我们：恒力将产生恒定的加速度。对于非常高速的情况这条规律不适用。因为一个物体的质量随速度增加而增大。在我们这个实验中，假定电梯的速度不是很大，因而，电梯在伽利略系中将有一个加速运动。

一个人待在电梯里让他的手绢和手表下落。电梯外伽利略系中的人认为手表和手绢会撞到地板上。这是由于地板因其加速度而向它们(手绢和手表)撞过来。事实上，电梯外的人将会发现表和地板以及手绢和地板间的距离以相同的速率在减小。另一方面，电梯里的人会注意到他的手表和手绢有相同的加速度，他会把这归因于引力场。

这两种解释看起来似乎一样：一边是一个加速运动，另一边是引力场，即加速运动与引力场引力效果是等效的。

让我们再做一个实验来证明引力场的存在，一束光通过窗户射在对面的墙上。我们的两位观察者是这样解释的：

在电梯外的人告诉我们：光通过窗户以恒定的速度(当然了!)沿一条直线水平地射进电梯，照在对面的墙上。但由于电梯正在向上运动，所以，光线的照射点应在此入射点稍下的位置上。

电梯里的人说：我们处于引力场中。由于光没有质量，它不会受引力场的影响，它会恰好落在入射点正对的点上。

两个观察者的意见不一致。然而在电梯里的人犯了个错误。他说光没有质量，但光有能量，而能量有一个质量($m=E/c^2$)，因此光将有一个向地板弯曲的轨迹，正像外部的观察者所说的那样。这种现象只能在非常强的引力场附近被观察到。这已经被证实：由于太阳的巨大质量，光线在靠近太阳时会发生弯曲。这个实验是爱因斯坦理论(广义相对论)的首次实证。

从所有这些实验中我们得出结论：通过引入一个引力场我们可以把一个加速系视为伽利略系。将其引申，我们认为它对所有的运动都适用，不论它们是旋转的（向心力被解释为引力场）还是不均匀加速运动（对不满足黎曼条件的引力场可以通过数学方法加以转换）。

**2. 广义相对论的新证据**

到目前为止只可能从广义相对论得出两个可以由观测检验的推论，即光线因太阳引力场而发生弯曲，以及来自遥远星球的光的谱线与在地球上以类似方式产生的（即由同一种原子产生的）相应光谱线比较，有位移现象发生。

2004 年 4 月 20 日发射上天的引力探测器一年间收集的地球引力场的资料堆积如山。如果爱因斯坦广义相对论是正确的，则由于地球的旋转，地球的引力场应被卷成螺旋形的。美国宇航局和斯坦福大学的科学家们分析了这些资料，证明确实存在这种引力涡旋。

**3. 广义相对论的意义**

相对论从逻辑思想上统一了经典物理学，使经典物理学成为一个完美的科学体系。狭义相对论在狭义相对性原理的基础上统一了牛顿力学和麦克斯韦电动力学两个体系。广义相对论又在广义协变的基础上，通过等效原理，建立了局域惯性场与普遍参考系数之间的关系，得到了所有物理规律的广义协变形式，并建立了广义协变的引力理论，而牛顿引力理论只是它的一级近似。这就从根本上解决了以前物理学只限于惯性系的问题。相对论严格地考察了时间、空间、物质和运动这些物理学的基本概念，给出了科学而系统的时间观和物质观，从而使物理学在逻辑上成为完美的科学体系。

广义相对论建立了完善的引力理论，而引力理论主要涉及的是天体。到现在，相对论宇宙学进一步发展。而引力波物理、致密天体物理和黑洞物理这些属于相对论天体物理学的分支科学都有一定的进展，吸引了许多科学家进行研究。

从广义相对论出发建立起来的引力理论是目前最好的一种引力理论。在现有的几种与广义相对论竞争的理论中，广义相对论占有明显的优势，不过它是不是唯一可能的正确理论，尚未有定论。所以，人们非常关心对广义相对论的实验检验，并且期望通过各种实验检验，进一步丰富和发展这一理论。

## 11.4 宇宙学与大爆炸理论

苍茫宇宙，星球无数，浩瀚的宇宙演绎着自然规律、法则和原理，启迪人类的智慧，演义着丰富多彩的文化和科学技术。千百年来，人们一直在追寻：宇宙是什么，宇宙到底有多大？宇宙的过去将来如何？物理学家更具体地追问：宇宙是如何演化的？

### 11.4.1 发展历史

**1. 关于静态宇宙的构想**

远在公元前 340 年，亚里士多德在他的著作《天论》（*On the Heavens*）中就已能提出两条有力的论据，以支持地球是圆球而不是平板的信念。第一，他认为月食是因地球运行到太阳与月亮之间而引起的。地球在月亮上的阴影总是圆的，这只有当地球呈球形时才有可能。

古希腊人认为地球必定是球形的，那就是为什么人们首先看到出现在地平线上的船帆，然后才看到船身。亚里士多德认为，地球是静止的，而太阳、月亮、行星和恒星都在环绕地球的圆形轨道上运行。他坚信这一点，因为出于神秘的原因，他觉得地球是宇宙的中心，而圆

周运动是最为完美的。

亚里士多德的构想在公元1世纪被托勒密加工成为一个完整的宇宙模型。地球位于中心,环绕着它的是八个天球,它们承载着月亮、太阳、恒星和当时所知的五颗行星:水星、金星、火星、木星和土星。为了解释观察到的它们在天空中相当复杂的轨迹,这些行星是各自沿其所在天球上较小的不动的恒星,它们总是互相保持不变的相对位置,但它们一起在天空旋转。在最外面的天球之外有什么,这是永远搞不清楚的,但它肯定不是人类可观察的宇宙的一部分。

托勒密(Claudius Ptolemaeus,英文 Ptolemy,公元90—168)的模型,为观测天体在天空中的位置,提供了一个合理的精确系统。但是,为了准确预测它们的位置,托勒密不得不做出一个假设:月亮运行的轨道在某些时候与地球的距离,是其他时候的两倍。而这意味着,月亮有时候看起来要比通常看到的大两倍。托勒密明白这是一个缺陷,不过他的模型虽然不是无人非议,却毕竟为大多数人所接受。它也被天主教会采纳为与《圣经》相符的对宇宙的描述。它最大的优点是,他把恒星天球之外的空间留给了天堂和地狱。

然而,1514年,一位波兰僧侣,哥白尼提出了一个简单得多的模型。起初,为了避免被指控为异端邪说,哥白尼匿名发表了他的模型。他的构想是这样的:太阳静止地居于中心,地球和其他行星在环绕太阳的圆形轨道上运行。对哥白尼来说可悲的是,几乎一个世纪之后,他的构想才得到认真的对待。这时有两位天文学家——德国的开普勒和意大利的伽利略——开始公开支持哥白尼理论,虽然它预测的轨道与观察到的不完全相符。

1609年,伽利略开始用刚刚发明的望远镜观察夜空。当伽利略观察木星时,他发现它被若干个小卫星,或者说是它的月亮包围着,它们环绕着它运转。这意味着,万物并非如亚里士多德和托勒密所设想的那样,都必须直接环绕地球运行。当然,尽管看到的是木星的卫星绕着木星转,但认为地球静止地居于宇宙中心,而这些卫星沿着极其复杂的轨道绕地球运行,仍然是可能的。不过,哥白尼的理论要简单得多。

与此同时,开普勒修正了哥白尼的理论,他指出行星不是沿圆周轨道运行,而是沿椭圆轨道运行。这样,预测终于与观察吻合了。就开普勒而言,椭圆轨道只是一个特定的假设——椭圆显然不如圆完美,因此这是一个相当讨厌的假设。虽然椭圆轨道与观察非常一致,但他无法将这个几乎偶然的发现与他的行星被磁力驱使环绕太阳运行的思想统一起来。

直到1687年,牛顿出版了他的《自然哲学的数学原理》,才给出了一种解释,牛顿不仅提出了物体如何在空间和时间中运动的理论,而且推演出分析这些运动所需的数学方法。此外,牛顿还作为公设提出了万有引力定律。

根据万有引力定律,牛顿进一步证明,引力使得月亮沿着椭圆轨道环绕地球运行,而且使得地球和各行星沿着椭圆轨道环绕太阳运行。哥白尼模型摆脱了托勒密的天球,与此同时也抛弃了宇宙有固定边界的思想。当地球绕着太阳转时,那些不动的恒星的相对位置看起来不会改变。因此,很自然地会得出这样的假设:这些恒星是与我们的太阳类似,但遥远得多的星体。

**2. 关于宇宙稳定性的构想**

20世纪以前,没有人曾提出过宇宙是膨胀或收缩的,这反映着那时的主流思潮。过去普遍认为,宇宙或者是永远以一种不变的状态存在,或者是在远古的有限时间内被创造成大体上如同我们今天看到的模样。这可能部分是由于人们有信仰永恒真理的倾向,以及从“人

有老死,宇宙永恒"的思想中他们得到慰藉。

有些学者意识到,牛顿的引力理论表明宇宙不可能是静态的,但就连他们也不想提出宇宙可能是膨胀的看法。相反,他们试图将引力在极远的距离处变成排斥力,以修正引力理论。这对于预测行星的运行没有明显的影响。但是,由于邻近恒星之间的吸引力被来自较远恒星的排斥力所抵消,因此这种修正允许无限分布的恒星保持平衡状态。

现在认为这种平衡是不稳定的。如果某一区域内的恒星互相之间稍微靠拢一点,它们之间的吸引力就会变得更强,从而超过排斥力。这就意味着这些恒星会继续相向坠落。反过来,如果这些恒星互相之间稍微离开远一点,则排斥力就会占上风,从而使它们离得更远。

哈勃在1929年的里程碑式的观察发现:无论向哪里看,遥远的恒星都在快速地离我们而去。换句话说,宇宙在膨胀。这意味着在较早的时候星体可能靠得比较近。实际上,似乎在大约100亿年或200亿年前的某一时刻,它们恰好都在同一位置。

这个发现最终将宇宙开端的问题带进了科学的领域。哈勃的观察启发科学家们进行假设与猜想:有一个可以称之为大爆炸的时刻,那时宇宙是无限小,因此其密度是无限大。并且进一步假设:如果有在此时刻之前发生的事件,那么它们对现在发生的一切没有影响。可以忽略它们的存在,因为它们不会有可观察到的后果。由于更早的时间根本无法定义,因此在这个意义上可以说,时间的开端就是大爆炸。

### 11.4.2 基本原理、理论结构

#### 1. 大爆炸模型的基本原理

大爆炸宇宙论的主要观点是:宇宙有过一段由热到冷、由密到稀不断膨胀的演化史,其过程犹如一次规模极其巨大的超级大爆炸。

根据这一学说,大爆炸时物质世界都集中在一个温度极高、密度极大的很小的范围。大爆炸开始后0.01秒,宇宙的温度约为10000亿度,其物质的主要成分为轻粒子(如光子、电子或中微子),质子和中子只占十亿分之一,所有这些粒子都处于热平衡状态。

由于整个体系在快速膨胀,温度很快下降。大爆炸后0.1秒,温度下降到300亿度,中子与质子之比从1下降到0.61。1秒钟后,温度下降到100亿度。随着密度减小,中微子不再处于热平衡状态,开始向外逃逸,电子-正电子对开始发生湮没反应。中子与质子之比进一步下降到0.3,但这时由于温度还太高,核力仍不足以把中子和质子束缚在一起。

大爆炸后13.8秒,温度降到30亿度,这时,质子和中子已可形成像氘、氦那样稳定的原子核。35分钟后,温度降到3亿度,核过程停止。但由于温度还很高,质子仍不能和电子结合形成中性原子。原子是在大爆炸后约30万年才开始形成的,这时的温度已跌到3000度,化学结合作用已足以将绝大部分自由电子束缚在中性原子内。

到这一阶段,宇宙的主要成分是气态物质,它们慢慢地凝聚成密度较高的气体云,再进一步形成各种恒星系统。这些恒星系统又经历了漫长的演化,成为人们今天所看到的宇宙,而星系中恒星的演化又产生了碳、氧、硅、铁等元素。为了解决大爆炸宇宙模型在解释宇宙极早期状态时所遇到的一些困难,20世纪80年代初,人们提出了暴胀宇宙论。这一学说对可观测宇宙的描述,除大爆炸发生后最初的$10^{-30}$秒外,与标准大爆炸模型符合得很好,但对最早那一瞬间的描述却大为不同。

根据暴胀模型,宇宙曾经历过一个极短时间的极快速膨胀(称为"暴胀"),在这段极短的时间内,物质处于称为"假真空"的奇特状态。由于假真空引起的引力排斥,宇宙按指数律加

速膨胀,其尺度每过 $10^{-34}$ 秒便增大一倍。在这种惊人的爆发式增长中,宇宙的所有质量和能量从完全的真空中产生出来。在这一新模型中,极早期宇宙的尺度比标准模型所认为的小得多,标准大爆炸模型中的一些困难在这里可以找到简单解释。

大爆炸实际上是从"奇点"开始发生的。大爆炸后的 $10^{-43}$ 秒被称为普朗克时期,在这一时期,可观测宇宙的尺度甚至比原子核还要小得多:半径只有 $10^{-23}$cm,密度却高达 $10^{90}$kg/cm$^3$,温度则高达 $10^{32}$ 度。

经过约 150 亿年的不断膨胀、冷却。目前可观测宇宙的范围已达 150 亿光年,其物质平均密度只有 $2\times10^{-31}$kg/cm$^3$,即每立方米空间只能分摊到一个氢原子。今天的宇宙与诞生极早期的宇宙相比,其尺度差异为 $10^{61}$ 量级,密度差异为 $10^{124}$ 量级,温度差异为 $10^{32}$ 量级。

**2. 大爆炸模型的理论结构**

随着支持暴胀理论的证据的积累,一个解决此问题的方案也提了出来。1998 年由帕尔马特(Saul Perlmutter,1959— )领导的一个小组和澳大利亚堪培拉附近的斯特朗洛山天文台的布赖恩·施米特(Brian Schmidt)用各自的技术测量了遥远星系中爆发的超新星到地球的距离。他们测量了爆发出光线的颜色,在绝大多数情况下,一个目标越远,看上去它就显得越红。在现代天文学中这种红移是一种使用非常流行的测距方法。天文学家已经知道超新星爆发有着十分相近的光度。所以,他们可以通过测量爆发所显现的亮度来推断其距离。他们的小组发现,通过亮度测量的超新星爆发的位置要比按照红移计算的距离远得多(红移基于宇宙膨胀速率正在减慢的假设)。这个发现使他们和其他天文学家提出了相反的观点:宇宙的膨胀正在加速,一些无法解释的神秘力量把宇宙向外拓展。

暗能量现象看上去很奇特。但是依照相对论理论,质量和能量是等价的。宇宙学家在思考产生这种神秘力量所需的能量时发现,这些能量恰好说明了宇宙图像上所欠缺的质量。如芝加哥大学的宇宙学家迈克尔·特纳(MichaeI Turner)说:"就像恰好吻合的齿轮,严丝合缝,十分有趣。人们很快就会接受这一看法。"

由美国国家航空航天局(NASA)的威尔金森各向异性微波探测器所得到的最新的宇宙微波背景数据做补充。现在的宇宙图景比以前清楚多了。总的来说,各种各样的宇宙微波背景研究已经证实宇宙的确是平坦的。威尔金森探测器给出了宇宙的一个组成清单:23%的暗物质,73% 的暗能量,剩下的 4%才属于星系、恒星和我们人类。

宇宙的年龄也被认定为 137 亿年,误差在 1%以内。全部物质的密度和暴胀理论所预言的相符,误差范围在 2% 以内。"这使宇宙学的可信度提高了一个大台阶,"费城的宾夕法尼亚大学的宇宙学家马克思·蒂格马克(Max Tegmark)说:"宇宙学已经变成了真实的、实在的科学。"

虽然这一巨大成就赋予了宇宙学理论前所未有的更可靠的立足点,但我们认识中的漏洞还是相当多的。天文学家确信暗物质和暗能量自始至终一直就存在,但不知道它们到底是什么,如何找到它们。就像特纳所说的"宇宙的 96%都是我们从未见过的东西"。

暗物质涉及的是更古老的、同时可能也是更简单的问题。理解它的关键在于它对恒星、星系的作用。按照广义相对论所说,任何质量都会使它周围的空间发生扭曲。当遥远物体发射出的光在暗物质近旁穿过时,它就会弯曲——这个作用叫作引力透镜。过去,天文学家只能探测能产生显著引力透镜效应的大的暗物质团块。但借助于计算机程序,他们现在可以探测到很微弱的扭曲效应。

美国夏威夷大学的天文学家尼古拉斯·凯泽(N-eholas Kaiser)说,这些引力透镜探测

计划正在进行中,在未来的几年内它们会有结果产出。他预言探测到的引力透镜数目将会以幂级数增长,这为天文学家测定当前宇宙的暗物质分布带来了希望。虽然这些探测可能对研究暗物质的本质没有多大的帮助,但它们会为暗物质在宇宙演化中扮演的角色提供更有力的约束。

对于暗物质是如何与其他物质相互作用的,宇宙学家也知道得很少。粒子运动的越快,它传递给任何与其相撞击的粒子的能量就越多。在早期宇宙中,如果暗物质是以接近光速的速度运动,它就会在物质成团、聚合,进而形成恒星和星系的过程中留下痕迹。天文学家能够看到宇宙中非常遥远之处恒星和星系的形成,但到目前为止他们还没有发现任何快速移动的暗物质与其他物质作用而留下的痕迹。

这使得许多宇宙学家推测暗物质可能由相对慢速运动的重粒子组成,而且这些重粒子很少与可见物质发生作用。这种推测引起了高能物理学家的兴趣,因为如果是这样,他们便可以为宇宙学家提供帮助。虽然现在粒子加速器还没有产生令人兴奋的成果,到 2007 年,欧洲核子中心(CREN)的科学家将有更强大的新设备—— 大强子对撞机投入使用(理论计算显示此对撞机可以产生暗物质的候选者),这会帮助宇宙学家建立更好的宇宙模型。

暗能量是更让人头疼的问题,问题的答案要依靠对真空性质的探寻。依据量子理论,即使在真空中,粒子和与其相对应的反粒子也在不断地产生和湮灭。一些科学家已在猜测这种真空能量可能就是加速宇宙扩张的力量。但理论预测的真空能量强度比今天探测到的暗能量强度要高 120 个数量级。

暴胀理论的创始者之一、普林斯顿大学的理论物理学家保罗·斯坦哈特(Paul Steinhart)表示,另一种观念认为暗能量可能是暴胀的孪生兄弟。如同暗能量一样,暴胀力反抗引力,把所有的东西向外推。但暴胀是当前向外作用力强度的好多倍,并且似乎在大爆炸后不久就消失了。理论学家现在正在用暴胀理论的弱态方程做实验来看他们可否描述暗能量。到目前为止,还没有出现可靠的结论。

**3. 大爆炸学说带来的困惑**

大爆炸宇宙论成功地解释了一些十分重要的观测事实,深得大多数天文学家的青睐。但是,从 20 世纪 80 年代后期开始,若干新观测结果的出现使这一理论受到不少挑战。持反对意见的学者对此提出了责难,而大爆炸宇宙论家则继续发展这一学说,以进一步对新的观测事实做出合理的解释。下面列举大爆炸学说遇到的一些挑战。

(1)“爆胀”意味着什么

在伽莫夫(G. Gamov,1904—1968)提出大爆炸之后,科学家们从各个角度研究并发现:标准宇宙模型给出的宇宙演化图景看上去很完美,可是也存在着一些难以解决的问题。

虽然从整体上来看是近似均匀的,但为什么从一个较小的局部看绝对不均匀(例如很多恒星组成星系)?为了弥补标准宇宙模型的这些缺陷,解释为什么出现了今天这样安定的宇宙,引入“爆胀”概念。“爆胀”从 $10^{-36}$ 秒持续到 $10^{-34}$ 秒,宇宙的尺度一下子膨胀到 1000 万千米,体积增大了 $10^{29}\sim10^{30}$ 倍,从而可用于解释今天的网状宇宙了。

为了解释宇宙初期的平坦性问题,必须要使用“爆胀”概念。但是其膨胀的速度远远超过了光速,用这样的速度,一秒钟就可以穿过可视宇宙(直径 300 亿光年)百万次之多。这无法用相对论来解释,故而又特别指出,此时此刻在宇宙开天辟地时相对论是失效的。

(2)宇宙的年龄

大爆炸宇宙学估计宇宙年龄为 100 亿～200 亿年,认为所有的恒星都是在原始爆炸后

温度下降时产生的。而至今人们观测的结果表明，各种天体的年龄没有发现超过200亿年的。

因此，似乎这证明了大爆炸产生了宇宙。在我们研究"脉冲星"的过程中，发现有一颗射电脉冲星G-P1953的年龄却为450亿岁，整整是现在流行的宇宙年龄的3倍还多。这等于说，在某个家庭中，有个孩子的年龄比其家长的年龄还大，这怎么可能呢？显然，是对宇宙的认识出了毛病！这里还没有谈及的类星体，它们的寿命，按哈勃定律计算是远远大于150亿年的星体。

(3) 元素丰度的疑问

大爆炸论提出的第三个证据是：宇宙中除了氢元素以外，最丰富的元素是氦。而且信奉大爆炸学说的科学家普遍认为：氦和其他几种轻元素是在宇宙早期3分钟内合成的，其他元素则在恒星演化的漫长过程中形成。具体地说就是宇宙温度降到了$10^{14}$K时，质子和中子结合形成氘或氦的核；当温度降到$10^{14}$K时，核形成停止，大约持续了3分钟，就认为全部物质质量的25%～30%变成氦核，而这正好与今天的氦丰度相符。

这个观测事实别的理论无法解释，而大爆炸宇宙学却能很好地解释。然而，事实上并不是这么回事，这是一种牵强附会的解释。因为大爆炸认为是它揭示了氢、氦等物质的形成过程，可是对同时形成的氘、锂等元素问题上却仍然不能给出解释。尽管有些天文学家设法自圆其说，他们认为氢、氦是宇宙大爆炸产生的，而氘、锂同位素是活动的星系核喷流出来的。其次，假如是"在3分钟内全部物质的25%～30%变成了氦核"，这个聚变核反应将放出多大的能量？我们可以大约算一下。

$$E=25\%M_{宇宙}c^2\times0.007=0.25\times10^{53}\,\text{kg}\times(3\times10^8\,\text{m/s})^2\times0.007=1.57\times10^{67}\,\text{J}$$

这个能量要比"创世"大爆炸的能量大得多了，难道又发生了第二次宇宙大爆炸？要知道，当年大爆炸的"点火"能量才是$10^{16}$(即$10^9$J)。

那么，氦是怎么产生的呢？我们知道，宇宙中的恒星像人一样，具有生、长、老、死，只不过时间是以天文数字来计算的。氦就是年轻时恒星内的核反应"产生"的。也就是说，恒星就是产生元素的"熔炉"，它的第一批产品就是氦元素。至于碳、氧、氮和其他更重的元素，也只能在恒星内部通过核反应产生(然后在这些恒星变成超新星爆炸时被抛射到星际云中)。大爆炸是产生不了这些元素的，包括氦在内。

考察一下我们的太阳。按照太阳上元素的组成情况，太阳绝不会是宇宙中形成的第一代恒星(假如有大爆炸的话)。因为太阳中含有只在超新星爆炸时才能产生的重元素。因此，很可能在太阳形成之前，在离太阳几十光年的地方，有一颗质量为太阳十几倍，甚至几十倍的恒星，变成了超新星，超新星爆炸时抛射出的碳、氧、氮等元素，污染了附近的一块星际云。超新星爆发在星际空间中形成的冲击波，使得这块星际云受到压缩，并使星际云破裂，其中的一个碎块，就是原始太阳星云，然后再演化成太阳。至于太阳为第几代恒星，计算表明，大约为第四、第五代恒星，也就是说太阳家族的历史有400亿～500亿年，而这与大爆炸的年龄150亿年绝对是矛盾的。

(4) "3K"辐射的探讨

"宇宙微波背景3K辐射"是"大爆炸派"自夸的主要观测事实之一(另一则是星系退行的哈勃定律)。

但事实上根据霍金的理论是得不出这一辐射的3K温度的。下面我们可按照霍金的公式计算一下，当大爆炸发生30万年时，温度已降为3000K。经过了150亿年后的今天，温度

应降为 $X$(K),则有等式:

$$3000(\text{K}) \times V_{20\text{万光年}} = X_K \times V_{150\text{亿光年}}$$

$$3000(\text{K}) \times \frac{4\pi}{3}(30\ \text{万光年})^3 = X_K \times \frac{4\pi}{3}(150\ \text{万光年})^3$$

$$3000(\text{K}) = X_K \times (5 \times 10^4)^3$$

$$X_K = 2.4 \times 10^{-11}(\text{K})$$

可见,霍金所说的 3K 辐射属于“张冠李戴”,就是退一万步说,假如 3K 是“余温”的话,那将表明,该辐射是来自曾发生爆炸的遥远的地方。而它经过了 100 亿年才到达地球的。换句话说,大爆炸位置离地球有 100 亿光年,如果是这样的话,那么在地球上,迎着爆炸的一面,将会看到星斗满空一切正常,可视距离为 100 亿光年。而背着爆炸的一面,将面对虚空。可视距离几乎为零。这怎么可能?

那么,3K 温度究竟是什么?“黑色背景”的空间温度与银河系的“白色背景”的空间温度绝不会是一样的。而“白色空间”的温度大约 3K。我们的太阳系在白色空间内,所以有 3K 的“辐射温度”。

(5) 真空不空——大爆炸的致命伤?

大量的实验和详细的理论分析表明,“真空”是一个具有一定物理性质和一定物理结构的客观实在。换句话说,真空也是物质的一种存在形态。1947 年,物理学家们进一步发现,真空不仅充满物质,而且还会由于外界电荷的影响,使真空的物质正负电荷偏满,这就是真空的极化。由此看来,真空也是物质。这样就存在一个问题,大爆炸的奇点包不包括真空?

如果我们的宇宙来自于一场大爆炸,那么可以肯定,我们所观测到的宇宙是有限的。大爆炸刚发生时它很小,大爆炸之后,它迅速膨胀,越来越大。但是,它在有限的时间内是不会膨胀到无限大的。

那么,在我们这个膨胀着的宇宙外边是什么呢?是不是存在一个一无所有的“超真空”?另外我们的宇宙包不包含“物质真空”?

宇宙的起源众说纷纭,直到现在还是研究的热点。

### 11.4.3 关于宇宙学的假设与猜想

2006 年 6 月 19 日,2006 年国际弦理论会议在北京人民大会堂开幕,霍金作题为《宇宙起源》的报告,其中霍金关于宇宙开端之前无时间的类比证明有启迪意义。霍金说:“时间,用纬度来测量,在南极处有一个开端,长期以来,人们说宇宙的开端是正常定律失效之处,所以宇宙不应该有开端。而现在,宇宙的开端由科学定律来制约,所以反对宇宙有开端的论证不再成立。爱因斯坦的广义相对论将时间和空间统一成时空,但是时间仍然和空间不同,它正像一个通道,要么有开端和终结,要么无限地伸展出去。然而当广义相对论和量子论相结合时,在极端情形下,时间可以像空间中另一方向那样行为,这意味和我们摆脱世界边缘的方法类似,假定宇宙的开端正如地球的南极,其纬度取时间的角色,宇宙就在南极作为一个起始点。”

霍金探询了在宇宙开端之前发生了什么等一系列问题,类似如在南极的南边没有任何东西的“地球证明”及是否是一个对称图形,还有对称破坏时间箭头,也是一个世界性难题,在《时间之箭》一书中,他认为除热力学外,牛顿力学、相对论力学、量子力学中的时间箭头都是对称的,而且超弦理论也不能避免。所以彭罗斯(Roger Penrose,1931— )也认为,倘若

有一种量子引力理论，能把相对论力学和量子力学统一起来，而如果没有解决时间箭头问题，也不算是成功的。

根据大爆炸理论，宇宙在空间和时间两个方面都是有起端的，即大爆炸瞬间。在这之前，无空间和时间概念可言，也就没有“之前”可言。尽管这一观念与人们日常生活中的感受格格不入，但根据大爆炸理论就是如此。只要承认这一点，那么人们就会问，宇宙的终极在哪儿？什么时候发生？会是一种什么状态？这就是宇宙的结局问题。大爆炸后的膨胀过程是一种引力和斥力之争，爆炸产生的动力是一种斥力，它使宇宙中的天体不断远离；天体间又存在万有引力，它会阻止天体远离，甚至力图使其互相靠近。引力的大小与天体的质量有关，因而大爆炸后宇宙的最终归宿是不断膨胀，还是最终会停止膨胀并反过来收缩变小，这完全取决于宇宙中物质密度的大小。

理论上存在某种临界密度。如果宇宙中物质的平均密度小于临界密度，宇宙就会一直膨胀下去，称为开宇宙：要是物质的平均密度大于临界密度，膨胀过程迟早会停下来，并随之出现收缩，称为闭宇宙。问题似乎变得很简单，但实则不然。理论计算得出的临界密度为$5\times10^{-30}\mathrm{gcm}^{-3}$。但要测定宇宙中物质平均密度就不那么容易了，星系间存在广袤的星系间空间，如果把目前所观测到的全部发光物质的质量平摊到整个宇宙空间，那么，平均密度就只有$2\times10^{-30}\mathrm{gcm}^{-3}$，远远低于上述临界密度。

然而，种种证据表明，宇宙中还存在着尚未观测到的所谓的暗物质，其数量可能远超过可见物质，这给平均密度的测定带来了很大的不确定因素。因此，宇宙的平均密度是否真的小于临界密度仍是一个有争议的问题。不过，就目前来看，开宇宙的可能性大一些。所谓开宇宙就是恒星演化到晚期，会把一部分物质（气体）抛入星际空间，而这些气体又可用来形成下一代恒星。这一过程会使气体越耗越少，以致最后再没有新的恒星可以形成。通过蒸发作用，有能量的粒子会从巨大的黑洞中逸出，并最终完全消失，宇宙将归于一片黑暗。这也许就是开宇宙末日到来时的景象，但它仍然在不断地、缓慢地膨胀着。所谓闭宇宙中，膨胀过程结束时间的早晚取决于宇宙平均密度的大小。如果假设平均密度是临界密度的2倍，那么根据一种简单的理论模型，经过400亿～500亿年后，当宇宙半径扩大到目前的2倍左右时，引力开始占上风，膨胀即告停止，而接下来宇宙便开始收缩。

以后的情况差不多就像一部宇宙影片放映结束后再倒放一样，大爆炸后宇宙中所发生的一切重大变化将会反演。收缩几百亿年后，宇宙的平均密度又大致回到目前的状态，不过，原来星系远离地球的退行运动将代之以向地球接近的运动。再过几十亿年，宇宙背景辐射会上升到400K，并继续上升，于是，宇宙变得非常炽热而又稠密，收缩也越来越快。

在坍缩过程中，星系会彼此并合，恒星间碰撞频繁。一旦宇宙温度上升到4000K，电子就从原子中游离出来；温度达到几百万度时，所有中子和质子从原子核中挣脱出来。很快，宇宙进入“大暴缩”阶段，一切物质和辐射极其迅速地被吞进一个密度无限高、空间无限小的区域，回复到大爆炸发生时的状态。

人们也许会认为，看来开宇宙的结局要比闭宇宙好一些，因为从理论上说，只要有星系及大黑洞存在，并且作旋转运动，技术高度发达的人类或地外文明总有办法汲取能量并继续生存下去；而一旦发生大暴缩，似乎一切都在劫难逃。但人们毕竟还有几百亿年时间去思考这一问题，应该相信科学技术的进步和人类的发展能力，而完全不必去作毫无意义的杞人之忧。迎接每一个美好的明天并为此做出自己的贡献，这才是人们应该做的。

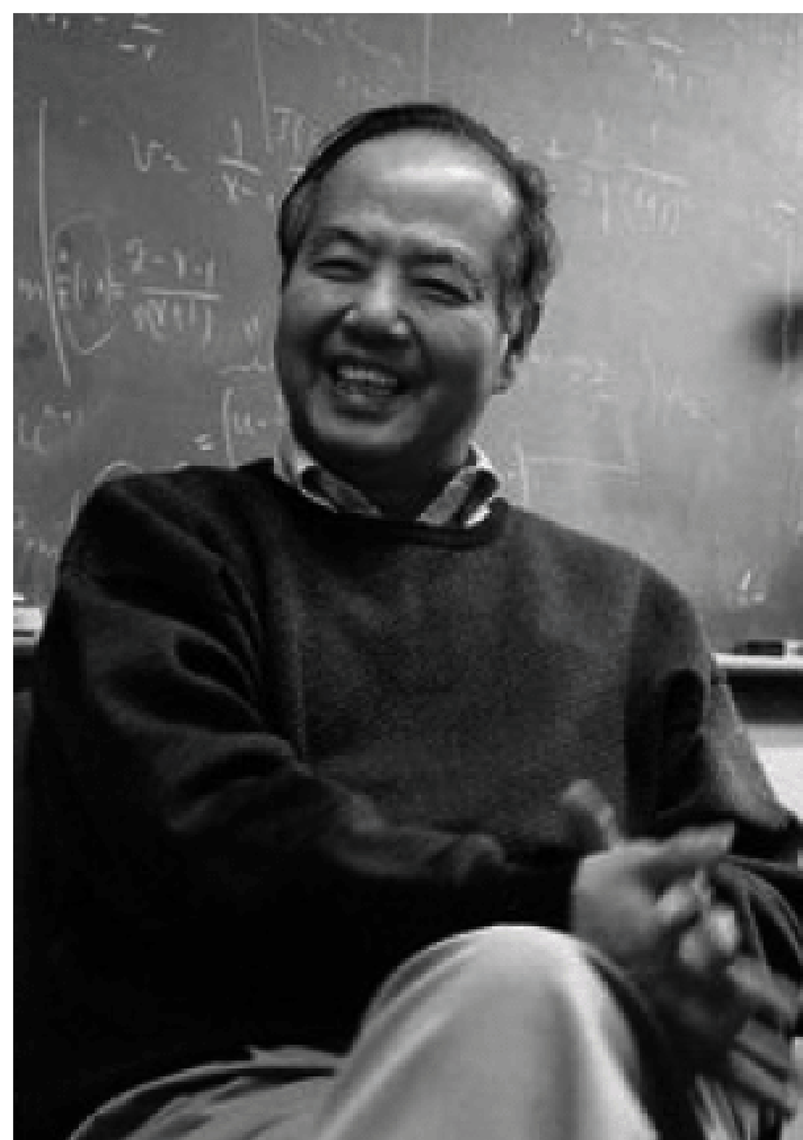
# 学习资源

## 一、中国近现代著名科学家(物理学)简介

杨振宁,1922年10月1日生,安徽省合肥县人。世界著名物理学家、诺贝尔物理学奖获得者。1949年,与恩利克·费米合作,提出基本粒子第一个复合模型;1956年与李政道合作,提出“弱相互作用中宇称不守恒理论”,与李政道共同获得了1957年诺贝尔物理学奖。杨振宁与李政道是最早两位获诺贝尔奖的中国人。杨振宁在统计物理、凝聚态物理、量子场论、数学物理等领域做出多项卓越的重大贡献。1997年紫金山天文台将其发现的一颗国际编号为3421号的小行星命名为“杨振宁星”。2017年2月,杨振宁放弃外国国籍成为中国公民,他的原中国科学院外籍院士正式转为中国科学院院士。2018年4月16日当选西湖大学校董会名誉主席。2019年9月21日,杨振宁被授予“求是终身成就奖”。目前,杨振宁担任清华大学高等研究院教授、香港中文大学博文讲座教授。



李政道,1926年11月24日生于上海,江苏苏州人,哥伦比亚大学全校级教授,美籍华裔物理学家,诺贝尔物理学奖获得者,因在宇称不守恒、李模型、相对论性重离子碰撞(RHIC)物理、和非拓扑孤立子场论等领域的贡献闻名。1957年,与杨振宁一起,因发现弱作用中宇称不守恒而获得诺贝尔物理学奖。1985年,他又倡导成立了中国博士后流动站和中国博士后科学基金会,并担任全国博士后管理委员会顾问和中国博士后科学基金会名誉理事长。1986年,他争取到意大利的经费,在中国科学院的支持下,创立了中国高等科学技术中心(CCAST)并担任主任。其后,成立了在浙江大学的浙江近代物理中心和在复旦大学的李政道实验物理中心。2018年4月7日,担任上海交通大学李政道研究所名誉所长。2004年任RIKEN-BNL研究中心名誉主任。2006年至今任北京大学高能物理研究中心主任。2016年获得“2015中华文化人物”荣誉。



吴健雄(1912.5.31—1997.2.16),生于江苏省苏州太仓浏河镇,美籍华人,其丈夫是华裔美国物理学家袁家骝,袁世凯次子袁克文的儿子。著名核物理学家,在β衰变研究领域具有世界性的贡献,被誉为“东方居里夫人”“核物理女王”“物理学第一夫人”。吴健雄是美国物理学会(APS)历史上第一位女性会长,也曾参与过曼哈顿计划,是世界最杰出的实验物理学家之一。


1928年的吴健雄

吴健雄于中华民国二十三年(1934年)从国立中央大学物理系毕业、获学士学位，于1940年从美国加州大学伯克利分校(UC Berkeley)毕业、获物理学博士学位。1952年，吴健雄任哥伦比亚大学副教授，1958年升为教授，1958当选为美国科学院院士，1975年获美国最高科学荣誉-国家科学勋章。1990年，中国科学院紫金山天文台将国际编号为2752号的小行星命名为"吴健雄星"，1994年吴健雄当选为中国科学院首批外籍院士。

吴健雄主要学术工作是用β衰变实验证明了在弱相互作用中的宇称不守恒，即用实验证明了核β衰变在矢量流守恒定律，结合μ子、介子和反质子物理方面的实验研究，从而验证"弱相互作用下的宇称不守恒"。该成果奠定了吴健雄作为世界一流实验物理学家的地位，许多著名科学家都为她没有因该项成就同杨振宁与李政道同获诺贝尔物理学奖而疑惑不平，但吴健雄已被公认为世界最杰出的物理学家之一。

1997年2月16日在纽约病逝，终年85岁。遵照她本人生前的愿望，吴健雄的骨灰安放在她的故乡中国江苏苏州太仓，纪念馆建于母校东南大学校园内。后来东南大学创建了吴健雄实验室和吴健雄学院以纪念这位杰出校友。

## 二、诺贝尔物理学奖简介

**1991年诺贝尔物理学奖**。把研究简单系统中有序现象的方法推广到更复杂的物质态，特别是在研究液晶和聚合物方面所作的贡献。获奖者：皮埃尔·德让纳(法国)。评价：德让纳的大名可能大家没听过，不过他有一个极为威风的名号"当代牛顿"。这是因为他的研究的影响涉及物理、化学、生物多个理论。他提出了一种"软物质"的概念，包括液晶、聚合物、胶体、膜、泡沫、颗粒物质、生命体系等。他把简单系统中有序现象的方法推广到比较复杂的物质形式，特别是推广到液晶和聚合物的研究中，从而获得诺贝尔奖。

**1992年诺贝尔物理学奖**。开发了多丝正比计数管。获奖者：G.夏帕克(法国)。评价：这次的荣誉属于那些为实验物理做出重大贡献的人，一般物理学需要记录粒子的轨迹。常用的记录方法是各种照相法，图片要靠特殊的测量工具进行分析，工作过程缓慢而又劳累。所使用的正比计数管确定粒子位置的精度大约是1厘米，不能满足粒子物理实验中记录粒子径迹的高精度和大面积覆盖的要求。夏帕克发明的多丝正比管解决了上述困难，得到的空间分辨率可达到1毫米或更小。20世纪70年代中期开始，他在多丝正比管的基础上发展了具有更高径迹定位精度的漂移室，进一步推动了粒子物理实验的发展。

**1993年诺贝尔物理学奖**。发现一对脉冲双星，为有关引力的研究提供了新的机会。获奖者：R.A.赫尔斯、J.H.泰勒(美国)。评价：要说脉冲双星就得先谈谈脉冲星，脉冲星的发现是60年代天文学四大发现之一，于1974年获奖。脉冲星其实是快速旋转的中子星(恒星演变的一种形态)，至于脉冲双星就是两颗互相环绕的脉冲星，他的发现可以通过计算轨道对爱因斯坦的广义相对论做出非常严密的实验证明。而且还有可能在计时方面提供比原子钟更准确的方法。

**1994年诺贝尔物理学奖**。在凝聚态物质的研究中发展了中子散射技术。获奖者：BN.布罗克豪斯(加拿大)、C.G.沙尔(美国)。评价：由于慢中子的大小和能量和凝聚态物质比较接近，而且可以和原子核直接发生作用(有些和电子作用)，所以，也被认为是探测凝聚态物质(把固态、液态和介于两者之间的各种状态，以及只有在低温下才存在的特殊量子态，还

包括稠密气体的物态统称为物质的凝聚态)的好材料。布罗克豪斯和沙尔各自独立开发的中子散射技术,对探测凝聚态物理学的发展起了促进作用,取得了重大成果。而凝聚态物理学由于应用前景广阔,实验技术成熟,可以说是当今物理学最热门的分支。

**1995年诺贝尔物理学奖**。检测到了中微子,发现了t轻子。获奖者:马丁·佩尔(美国)、弗雷德里克·莱因斯(美国)。评价:中微子的发现其意义我在前面讲过,不过佩尔才是第一个发现中微子的人,而1988年获奖几位是发现的另外一种中微子。除了中微子物理学还有一个轻子的概念,轻子就是不参与强相互作用的有半整数自旋(如1/2,3/2,5/2等)的粒子;轻子都是物质的基本粒子(从现在的观察水平看)而t轻子则是人类发现的第三种轻子。而轻子、中微子和夸克是人类目前发现的三类基本粒子,可以说是这次的发现是基本粒子领域的又一重大进展。

**1996年诺贝尔物理学奖**。发现氦-3中的超流动性。获奖者:戴维·李、奥谢罗夫、R. C.里查森(美国)。评价:关于氦的超流动性前面已经提过,是由卡皮查发现并由朗道提供解释。不过根据朗道的理论氦的一种同位素氦-3是不具备超流动性,但这与超导理论中的BCS(1972)理论矛盾。而上述三位在实验中发现了氦-3的超流动性解决了这一争端,并且还可以通过此来验证理论物理的一些其他问题。不过氦-3看似遥远不过很有可能在将来成为我们日常生活中的必需品。因为它是实现可控核聚变的重要材料。将来能源问题的一个解决方案。由于此物质在地球含量低而在月球含量大,所以这也是各国探月计划的一个重点。

**1997年诺贝尔物理学奖**。发明了用激光冷却和俘获原子的方法。获奖者:朱棣文(美籍华人)、W. D.菲利普斯(美国)、C.科昂-塔努吉(法国)。评价:利用激光冷却并捕捉原子的方法,犹如以喷水的方式来使一个行进当中的小球静止下来。通过用大量光子撞击原子,控制原子的速度,使原子尽量静止从而使观察更容易。激光冷却是冷却技术中产生温度最低的一种技术。

**1998年诺贝尔物理学奖**。发现了分数量子霍尔效应。获奖者:劳克林(美国)、斯特默(美国)、崔琦(美籍华人)。评价:分数量子霍尔效应就是在作量子霍尔效应的实验时在更低的温度和更强的磁场对霍尔效应进行研究,发现量子数的间隔在冯·克利青的基础上会出现更小的分立值。而斯特默教授和崔琦教授在1982年做出此发现。而劳克林(laughing)一年后对此做出了解释。

**1999年诺贝尔物理学奖**。阐明了物理中电镀弱交互作用的定量结构。获奖者:H.霍夫特(荷兰)、M.韦尔特曼(荷兰)。评价:电弱统一理论是20世纪最伟大的科学成就之一,而关于此理论有几个重要问题需要解决。格拉肖、温伯格和萨拉姆三人(1979)基本上解决了大部分问题。但仍有几个遗留问题需要解决。这两位对这一理论补上了临门一脚。使这一理论变得更加严谨,完整。

**2000年诺贝尔物理学奖**。发明快速晶体管、激光二极管和集成电路。获奖者:阿尔费罗夫(俄罗斯)、基尔比(美国)、克勒默(美国)。评价:这次的发现一半是奖励基尔比关于集成电路,开启了集成电路的时代。基尔比所属的得克萨斯州仪器公司也一直是这方面的强者,不过令他比较郁闷的是他隔了40多年才获奖,而发明晶体三极管的人则只隔了9年,奖金另一半则表彰另外两人对半导体基础元件有重大的发明和改进。

## 思考与练习

1. 迈克耳孙-莫雷实验的意义。
2. 时间延缓是绝对的还是相对的？
3. 孪生子佯缪是真实的绝对生物效应，还是相对的物理效应？
4. 谈谈孪生子效应与孪生子佯缪的区别与联系。
5. 谈谈广义相对论与狭义相对论的区别与联系。
6. 谈谈发现引力波的实验原理。
7. 大爆炸理论的实验依据是什么？
8. 谈谈对“宇宙年龄”推算的认识。

# 第12章

# 非线性物理世界

所谓线性，通俗地讲，描述某系统两个变量之间的正比例关系就是线性关系；从自然系统的因果关系的数学描述上来讲，是指系统的模型方程的解满足线性叠加原理，即方程的任意两个解的线性组合仍然是该方程的一个解。对于一个实际的非线性系统，哪怕一个小扰动，比如，初始条件发生一个微小改变，都可能造成系统在往后时刻行为的巨大差异。叠加原理的失效也将导致傅里叶变换方法不适用于非线性系统的分析，所以，对于非线性系统演化行为的解析研究是相当困难的。

20 世纪初，彭加勒首先意识到在确定论系统中有混沌现象的存在，对牛顿严格的决定论提出了质疑，由此拉开了非线性科学研究的序幕。非线性科学是一门研究复杂性的新科学。随着非线性科学的深入发展，人类逐渐认识到非线性并非只存在于宏观层次，微观领域和宇观领域本质上也是非线性的，即整个宇宙本质上就是非线性的。目前非线性科学研究的主要内容有以下几个方面：非线性映射的宏观特性、混沌与分形、动力学系统的时间反演问题、自组织与耗散结构、复杂性探索等。一般认为，混沌、分形和独立子是非线性科学的主题，并且它们三者是彼此联系的。

非线性物理学作为科学的一个新分支，如同量子力学和相对论一样，给了我们一个全新的思想。目前发展起来的非线性物理学科包括：非线性基础物理学（非线性光学、非线性声学、非线性动力学）、耗散结构理论、分形理论和混沌理论等。

## 12.1　熵增原理与耗散结构

熵增原理来源于热力学第二定律，经过克劳修斯和玻耳兹曼的研究以及申农在信息论里对熵概念的拓展，使得熵理论得到了空前的发展。熵增原理是被人们所接受的，但是根据熵增原理必然会导致热寂学的产生。科学家对此作了大量的研究，最后，普利高津在前人研究的基础上，创造性地提出了耗散结构理论。这一理论改变了人们对传统热力学的看法，将热力学的研究带入更广阔的前景。同时，随着耗散结构理论的产生与发展，其理论已经广泛运用于各个学科领域，为科学的发展，人类社会的进步做出了巨大的贡献。

### 12.1.1　熵理论的发展

熵理论的发展大致经历了平衡态的熵理论、非平衡态的熵理论和广义熵理论三个阶段。

**1. 平衡态的熵理论**

我们知道，熵理论的发展主要是熵概念的定义以及熵增原理的提出，这其中又经过了卡诺、克劳修斯等物理学家的研究，使得熵的内容得到了极大的丰富。卡诺在 1824 年创造性地提出了“可逆过程”这一基本概念，这是后来的熵和热力学第一定律的萌芽。

早在 17、18 世纪，人们都想造出永动机来，可是经过无数科学家的尝试都以失败而告

终。这引发了人们的思考，是否是人类忽略了某些自然界已有的、且未知的规律？这样的思考在后来直接导致了热力学第一、第二定律的发现。

克劳修斯在 1854 年，用热量与绝对温度的比值来表示熵，他根据卡诺定理得出，对于任意一个系统在可逆循环中，两个状态下所交换的热量与在这个过程中系统的绝对温度的比值的环路积分等于零；任意系统在不可逆循环中，两个状态下所交换的热量与在这个过程中的系统的绝对温度的比值不小于零。即：

$$\oint \frac{\mathrm{d}Q}{T} \geqslant 0 \tag{12-1-1}$$

根据式(12-1-1)，对于可逆循环则如图 12-1-1 所示，有

$$S - S_0 = \int_a^b \frac{\mathrm{d}Q}{T} \tag{12-1-2}$$

式(12-1-2)中的 $a$ 和 $b$ 是系统任意的两个平衡态，$S$ 是系统处于 $b$ 平衡态时的熵，$S_0$ 是系统处于 $a$ 平衡态时的熵。根据上式，可以看出，熵变只与始末两个平衡态有关，与过程无关。无论是沿着路径 1 还是沿着路径 2，得到的系统的熵变是相同的。

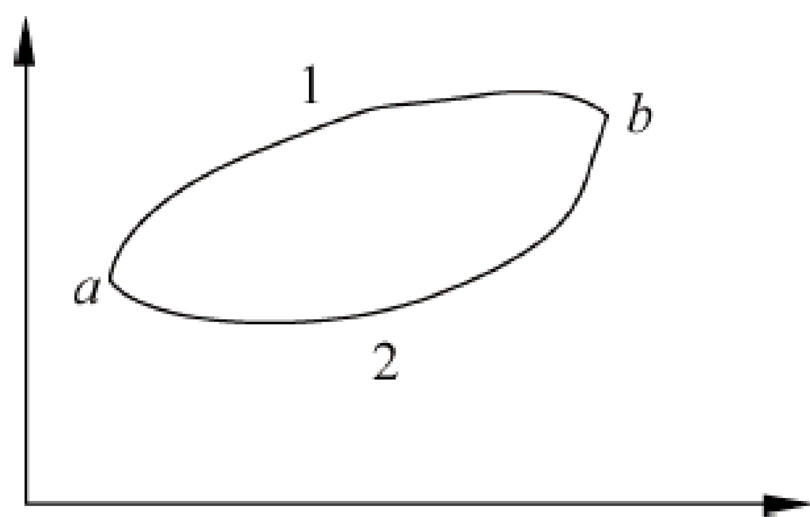


图 12-1-1　系统循环

对于不可逆过程，系统的熵有两个部分组成，一部分是系统在与外界环境相互作用过程中流进或流出的熵，称为熵流项，用 $\mathrm{d}_e S$ 表示；另一部分则是由系统的不可逆过程产生的，称为熵生项，用 $\mathrm{d}_i S$ 表示，于是可以得出

$$\mathrm{d}S = \mathrm{d}_e S + \mathrm{d}_i S \tag{12-1-3}$$

热力学第二定律描述的是任何一个自发的宏观热运动过程都是不可逆的。由此可以判断，对于一个孤立体系来说，如果它与外界没有联系，在该体系中所进行的不可逆过程的结果，是不可能凭借体系自身的任何过程而自动复原的；如果借助外界使体系复原，那么必然会对外界产生不可消除的影响。也就是说，一个孤立系统的熵只能增加，不能减少，只有在不可逆过程下才能达到极大值，这时系统处于平衡态。因为，孤立系统不可能从外界吸收物质或能量，那么在系统内部所发生的过程必定是一个封闭的过程，其熵只能增加，不能减少，直到系统达到平衡态，此时系统的熵具有极大值。由此可见，平衡态的熵理论的研究已经达到很高的水准了。

**2. 非平衡态熵理论**

20 世纪 30 年代以后，熵理论的发展突破了平衡态领域，发展到了非平衡态领域，这是熵理论发展的第二个阶段。这个过程首先是从昂赛格的倒易关系和普利高津的最小熵原理开始的，后来逐渐发展到了非平衡态的熵理论。

昂赛格提出的倒易关系是各种不可逆现象联系的元过程的时间反演不变性的严格结果，这成为熵理论研究的转折点，人们对熵的研究从平衡态转入了非平衡态。昂赛格也凭此获得了 1968 年的诺贝尔物理学奖。

熵理论的研究从平衡态到非平衡态的最大的困难便是对一个热力状态的描述问题。如果抛弃原有的状态变量和它们之间的关系的话，那么之前的研究便失去了意义。这个问题的解决得益于普利高津的“局部平衡假设”的提出。他在 1945 年发现了最小熵原理，随后他又对熵理论做了大量的研究提出了耗散结构理论等。至此，非平衡态下的熵理论研究也取得了长足的发展。

**3. 信息熵**

信息论中信息熵的引入，是熵理论发展的第三阶段。信息论所说的信息并非传统的信

息,信息熵的概念也比之前具有更丰富的内容和普适性,因此可以说信息熵是一种广义熵。玻耳兹曼曾经说过:“熵是一个系统失去信息的度量”,这是信息熵产生的前提。申农在1948年提出了信息量的概念,他将其与统计力学的熵联系起来,提出了“不确定性”这一概念,后经过布里渊的提议改为信息熵,从而解决了定量描述信息的难题。信息熵理论还在不断的发展中,其理论几乎已经运用到了所有领域。

信息熵的提出也是由热力学熵的启发得来的,著名的数学家冯·诺依曼建议将申农所提出的不确定度定义为熵。在一个事件中,我们得到的信息或提示越多,该事件的不确定性就越小。获得信息或提示过程中所降低的不确定度或减少的熵就称为负熵,也可以称为信息量,其公式为

$$I = -H = -\log_2 W \tag{12-1-4}$$

信息量 $I$ 表示的是一个体系的有序程度、复杂性和发展程度等。获得的信息量越大,体系的不确定度就越小,这与熵的性质刚好相反,因此将其称为负熵。

由以上内容我们可以将熵分为三类,即物理学中的熵:用来描述系统的混乱程度;分子运动论中的熵:用来描述分子运动的无序程度;信息论中的熵:用来描述信息论中的不确定状态。

### 12.1.2 对熵增原理的认识

热力学第一定律阐明了世界上的能量是守恒的,而热力学第二定律则说明,不可能将吸收的热量全部转化为功,说明有部分能量转化为了不可利用的能量。

**1. 关于热力学第二定律**

关于热力学第二定律的表述有两种,一种是克劳修斯表述:不可能把热量从低温物体传到高温物体而不引起其他变化;另一种是开尔文表述:不可能从单一热源吸收热量而使之完全转化为有用功而不引起其他变化。这两种表述都说明了一个问题,即与热现象有关的宏观热运动过程是不可逆的,具有时间上的单向性。

热力学第二定律的数学表达式为

$$\mathrm{d}S \geqslant \frac{\mathrm{d}Q}{T} \tag{12-1-5}$$

当 $\mathrm{d}Q=0$ 时,则 $\mathrm{d}S\geqslant 0$,这就是熵增原理。

式(12-1-5)可以表述为:在一个孤立系统内,系统的熵不减小。当这个过程可逆时熵不变;当此过程不可逆时熵增加。且整个过程朝着熵增大的方向进行,直到达到平衡态,此时熵最大。

**2. 熵增原理**

从熵增原理可以看出:从单一热源吸收热量而使之完全转化为有用功而不引起其他变化的过程是不能实现的。这就是由熵增原理导出的热力学第二定律的开尔文表述。由此可见,熵增原理就是热力学第二定律的另一种表述。

(1) 微观领域的熵增原理

平衡态体系内的分子的无规则运动要比非平衡态体系内分子的无规则运动更加复杂更加无序。因为,这种体系内的分子的运动,最无规则最混乱,才能使得体系内的各个地方的温度、压强、密度等达到一致,并且最后达到平衡态。因此,熵增过程就是孤立体系从非平衡态到平衡态的过程,就是体系内的分子的运动由不太乱到非常混乱的过程。由此可见,熵增原理在微观领域的意义就是体系内分子作无规则运动的混乱程度的量度。熵增加就是从有

序向无序发展的过程。

(2) 宏观领域的熵增原理

能量都是从非均匀分布向均匀分布转化而做功的。要想让能量转化为可以利用的能，就是要让能量用于做功。而要让能量做功就需要存在能量密度差，使能量从高密度区流向低密度区，在这个过程中便产生了功。从宏观角度来说，能量分布越不均匀，越混乱，则体系的熵就越小，其将能量转化为有用功的效率就越高。如果体系内的能量分布已经均匀，那么就没有能量差，更不会产生能流，也就不会获得有用功。熵增原理在宏观领域主要表现为不可利用能的增加和可用能的减少。时间具有单方向性，因此能量只能从有用能转化为无用能。

(3) 熵增原理是自然界最普遍的规律

自然界的所有规律都遵循熵增原理。无论是动物的生长发育还是植物的光合作用都是如此，植物的生长需要消耗其他物质的化学能，在光合作用的帮助下吸收太阳能做功，在这个过程中有外界的能量进入了植物中，而不再是孤立系统中发生的过程，这个过程遵循热力学第二定律。在植物生长过程中的能量来源是太阳，说明在开放系统中可以发生熵减小的过程，但是它是以周围环境熵增大为代价的。

目前随着人类社会科技的不断发展，经济也越来越发达，但随之而来的各种环境问题也越来越严重，如气候变暖、臭氧层空洞、冰川融化、环境污染等。这些问题都与人类的活动有关，也可以说这些问题都是熵的变化，熵在朝着增大的方向运动。而人类的目的是要用负熵来减缓这个熵增发生的过程。如今的世界是科学的世界，因此我们需要用科学的方法合理利用资源。所谓可持续发展，就是不断地为自然系统引入负熵流，维持自然系统的稳定性，使其熵增加的速度减缓。

**3. 对热寂学的批判**

所谓的热寂学是将熵增原理推广至整个宇宙而得到的“热寂”状态。该理论认为：宇宙越是接近于熵最大的极限状态，它继续发生变化的可能性就越小。当宇宙达到这个状态时，就不再会出现进一步的发展，宇宙将永远处于一种惰性的死寂状态。一百多年来人们为热寂学所困扰，而熵增原理也似乎证明了这个学说的可能性。根据熵增原理可知，每当我们消耗一定的燃料，如煤、石油等，都在增加宇宙的熵，也就是说每次活动都降低了能量的品质，从实际情况来看也确实如此。可是宇宙并没有变得很热，相比之下宇宙的温度要比地球低得多，也就是说熵并没有趋于最概然分布。宇宙也没有变得越来越无序，相反，宇宙却是生机勃勃。恩格斯也曾说过：放射到太空的热一定有可能通过某种途径转变为一种运动形式，在这种运动形式下，热能够重新集结起来。目前有研究表明，引力可以是熵减的，也就是说宇宙只要不无限膨胀，那么在宇宙中就会存在温差。

普利高津的耗散结构理论对此可以给出合理的解释，封闭系统（我们的宇宙）的能量是守恒的，系统可以通过与外界的物质与能量的交换实现熵减小的过程。

### 12.1.3 耗散结构理论

贝索曾多次向爱因斯坦提出时间及其不可逆性的问题，但是并没有得到很好的解答。随后又出现了“时间悖论”，这些都促使普利高津去建立一个新的理论体系。1945 年，他在原有的理论基础上又提出了最小熵原理。这个原理和昂赛格的倒易关系为近平衡区线性热力学的发展奠定了基础。随后，普利高津进行了大量的研究，他想将这一原理推广到远离平衡态的非线性系统，但是一直都没有成功。后来，他将远离平衡态系统和近平衡态系统做了

严格的区分并重新考察远离平衡态的系统的情况。1969年,他在一次国际会议上提出了他的理论,随后他又在自己所著的书中详细地阐述了这一理论。最终通过几十年的努力,他与布鲁塞尔的同事们创立了一种新的关于非平衡态的系统自组织理论——耗散结构理论。

耗散结构理论即用热力学和统计物理的方法,研究耗散结构形成的机理、规律和条件的理论。它是一个远离平衡态的非线性的开放系统,通过不断与外界交换物质与能量,使得系统内某个变量发生变化,当这种变化达到一定的阈值时,通过整个系统的涨落,系统可能会发生突变,由原来的混沌无序的状态转变成一种在时间和空间以及功能上有序的状态。耗散结构作为一种揭示复杂系统中自组织运动规律的新兴理论,为推动其他领域的科学发展起到了巨大的作用。普利高津的耗散结构的形成与之前的研究是分不开的,其理论的建立大致经历了以下三个阶段的研究。

**1. 开放系统**

我们知道,对于一个开放系统,其熵变是由两个部分组成的。一部分是系统在不可逆过程中系统产生的熵 $d_iS$,这一项是一个非负数(过程可逆是等于零);另一部分则是该系统与外界进行物质和能量交换时流入的熵 $d_eS$,这一项可以大于、等于、小于零,这主要取决于外界对系统的影响。如果外界流入的熵小于零,且远小于系统产生的熵,即

$$d_eS < 0, \quad 且\ |d_eS| > d_iS \tag{12-1-6}$$

则,系统的总熵变等于这两部分之和便小于零。这就意味着,在外界的影响下,通过不断地从外界获得物质或能量,系统可以由混乱状态变得更加有序。对于开放系统的研究是形成耗散结构的前提。

**2. 偏离平衡态的系统**

当系统在外界的影响下,使得作用与效果之间存在简单的线性关系(昂赛格倒易关系),这样的系统便偏离了平衡态。偏离平衡态的系统也是一个稳定的状态,称为非线性平衡定态。普利高津对线性非平衡态做了大量研究提出了最小熵原理,即系统在偏离平衡态的范围较小,且外部的控制条件恒定的情况下,系统的自发过程产生的熵具有最小值。

最小熵原理表明,偏离平衡态的系统总要选择一个最小能耗的状态,在这个定态下,系统产生的熵是最小的。这个状态是稳定的,即使受到外界的干扰,系统会朝着熵增加变大的状态转变,但是根据最小熵原理,系统还会回到这个定态。

处于非平衡定态的系统,如果撤销外界的干扰,系统必然熵增大,最终达到平衡态。达到平衡态时,系统的熵达到极大值。因此可以说,平衡态是非平衡定态的一个特殊情况。这种非平衡的定态的维持正是耗散了外界的可控的物质和能量的系统。对于偏离平衡态的系统的研究为后来耗散结构理论的提出起到了关键性作用。

**3. 远离平衡态的系统**

远离平衡态的系统,是指系统受到了外界强烈的作用,且作用与效果之间的关系并非线性关系。非线性非平衡态是热力学研究的最新阶段,研究这类系统,需要借助系统的动力学和热力学两种方法来解决。这类问题的解决相当复杂,普利高津在对前人研究开放系统和偏离平衡态系统的基础上,开创性地提出了耗散结构理论。

### 12.1.4 耗散结构的形成条件

耗散结构的形成是有很多条件的,根据耗散结构理论的解释可以看出,要想形成一个耗散结构,该系统必须是一个开放的系统。其次,系统必须是远离平衡态的,同时系统要靠涨落来达到新的有序状态。最后,系统内的诸多要素之间的相互作用必须是非线性的。

**1. 系统是开放的**

开放系统是能够自由地与外界发生物质能量交换的系统。只有开放的系统才能在与外界进行物质和能量的交换中获得负熵，使系统由混乱状态向有序状态发展。若系统是孤立系统，则与外界失去了联系，外界无法对系统产生影响，在这样的条件下，系统所自发的演化必然是朝着熵增大的方向，系统将会变得更加无序，自然也就无法形成具有动态平衡的有序的耗散结构。

普利高津在不违反热力学第二定律的前提下，引入了负熵流用来抵消系统自发过程中产生的熵增，使系统的总熵减小，进而使系统变得更加有序，从而解决了克劳修斯和达尔文的进化论的矛盾。他将热力学与进化论统一起来，使得物理学的规律与生物学规律有机结合，推动了科学的发展。

**2. 系统是远离平衡态的**

普利高津认为，开放的系统是耗散结构形成的必要条件，但不是充分条件。光有开放的系统而没有远离平衡态的话，是不会形成这种结构的。因为，如果系统处于平衡态，或者偏离平衡态的非平衡线性定态的话，系统是处于稳定状态或者趋于稳定状态的，这样的系统总是趋于平衡的。

处于平衡态的系统是稳定的，它的存在不需要依赖于外界就可以存在，而耗散结构是一种动态的平衡，其存在是依赖于与外界的物质能量的交换。耗散结构是一种非平衡态，也正是这种非平衡态，才使得系统有了与外界交换物质与能量的驱动力。因此，要想形成一个耗散结构，就必须通过外界，使一个开放的系统远离平衡态和近平衡态，使其产生交换物质与能量的趋势。

**3. 涨落是系统形成新的有序的驱动力**

涨落是系统的某个变量相对于平均数的偏离。涨落在不同的条件下起着完全不同的作用，在近平衡区，正是由于系统的涨落引起了系统运动的混乱，使系统变得无序的。而在非平衡系统中，涨落具有重要的积极作用，系统通过涨落从不稳定状态转化为新的稳定状态。当系统处于远离平衡状态时，系统内某个子系统的涨落可能导致整个系统的涨落，使整个系统发生突变，由混乱状态变成新的有序状态。耗散结构的实质就是子系统的涨落被放大，并且通过与外界的物质和能量的交换使系统最终趋于稳定。因此，涨落是系统由原来的平衡态演化到耗散结构的驱动力。

如图 12-1-2 所示，普利高津用图示表明了结构、功能和涨落之间的关系。这三者相互作用导致了系统新的有序的出现。因此，涨落是形成新的有序的驱动力。

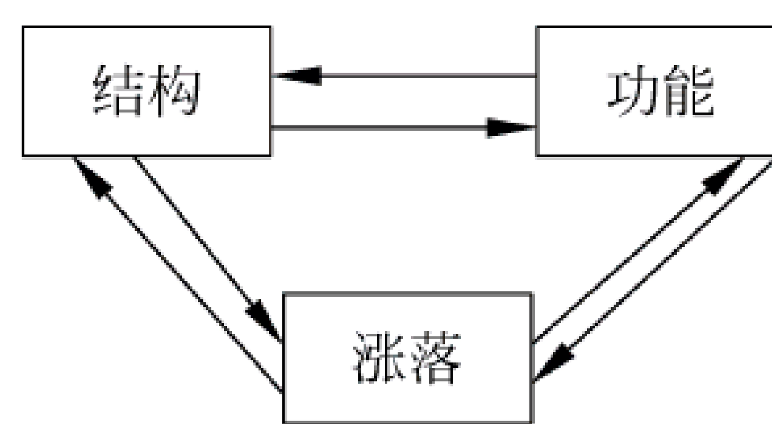


图 12-1-2 系统的结构、功能和涨落的关系

**4. 远离平衡态的系统的各要素之间存在着非线性的相互作用**

普利高津的耗散结构理论认为，对于一个近平衡态系统，其演化的结果只有两个，要么消亡要么无限增长。近平衡态系统的定态和非平衡约束是对应的，也就是说它只有热力学分支。即使通过外界条件可以将近平衡态系统转化为远离平衡态系统，因为作为扰动发展归宿的其他稳定分支不存在，所以在不稳定热力学分支附近的扰动也只能无限发展。

然而，远离平衡态的系统的演化相比于近平衡态是具有多种可能性的。对于近平衡态的系统而言，其非线性项与线性项相比影响是可以忽略不计的，因此只存在热力学分支的解；当非平衡约束把系统转化到远离平衡态时，非线性项就发挥了主要作用，一个非平衡约

束就可以对应多个定态解。如图 12-1-3 所示,在这些解当中,有的分支是稳定的,有的分支是不稳定的,随着时间的推移,系统最终会演变到其中一个稳定的分支上。由此可见,远离平衡态的系统的各要素之间的非线性相互作用也是形成耗散结构所必不可少的条件。

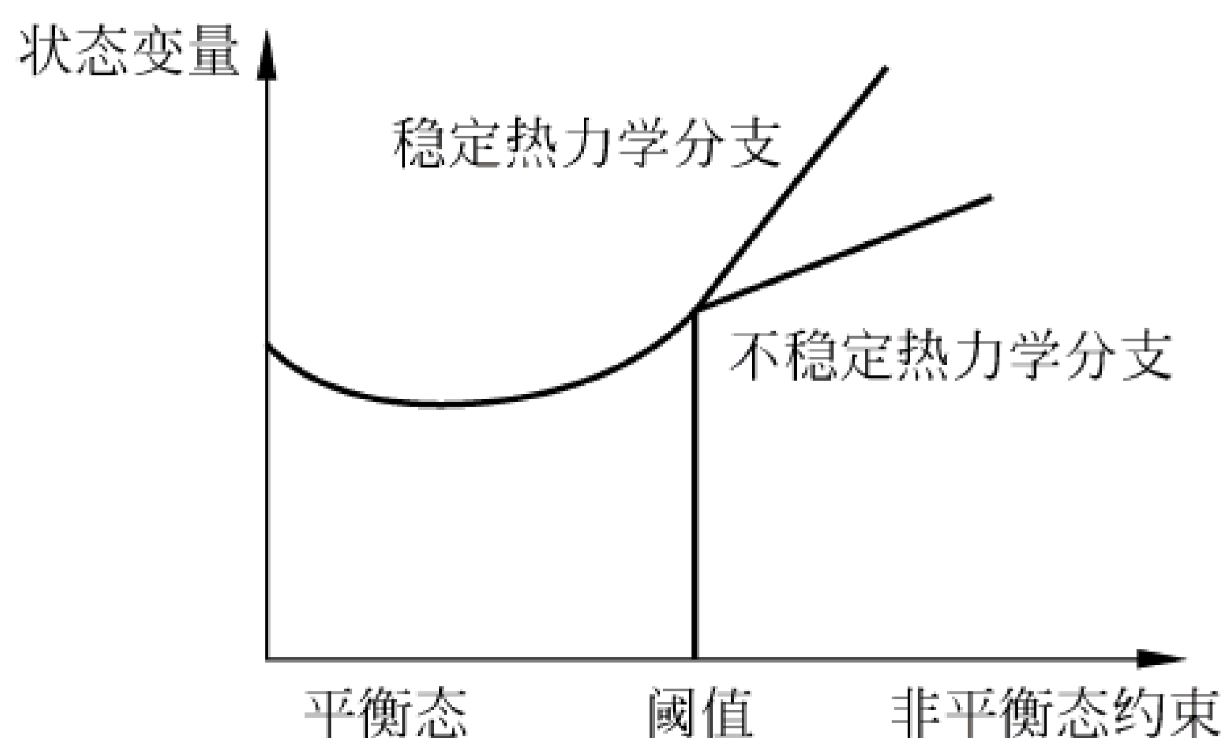



图 12-1-3 系统演化图

## 12.1.5 对耗散结构理论的认识

自组织过程是耗散结构形成的机制。复杂的系统在远离平衡态时,可能会出现巨型涨落,引发系统的一系列变化。在变化的分叉点前,系统通过对多种分支的随机选择,形成新的结构和状态。

**1. 耗散结构理论中的自组织过程**

所谓复杂的系统是指系统包含了较多相互作用的子系统和结构,同时也指系统与外部环境的相互作用。远离平衡态是指系统对于外部和系统内的细微变化变得异常敏感。通常情况下,这些影响对平衡态和近平衡态的系统的影响可以忽略不计,但是对于远离平衡态的系统来说,这种细微的扰动可能会使系统发生质的变化。在平衡态和近平衡态的系统中,小的涨落是对平均值的修正,系统最终还是会趋于平衡。而在远离平衡态的系统中,小的涨落可能不会衰减,而是被放大,这种巨型的涨落使系统发生剧烈的变化。当系统超过了热力学稳定的阈值以后,系统便到达变化的分叉点,即系统对涨落的反应会在几个稳定或不稳定的状态间进行随机选择。

在复杂变化中,一级分叉还不足以把所有可能的状态都包含在内,系统会通过逐级分叉形成新的状态。在这一系列的变化中,系统需要经历很多的稳定和不稳定过程。其中在两个分叉点之间的区域是稳定的,系统遵循物理学规律,如平衡态和近平衡态的热力学理论。而系统处于分叉点附近是不稳定的,这时候主要是靠系统的涨落和随机选择起作用。

**2. 耗散结构理论的普适性**

耗散结构不仅存在于自然界中,也存在于社会生活的各个领域。长期以来,人类在研究自然界和人类社会时,都是独立研究的,前者称为自然科学,后者称为社会科学。普利高津的耗散结构理论使它们得到了有机的统一。他运用了稳定的数学分析方法解决了体系有序演化的问题。其理论在物理学、化学、生物学、人口系统、环境系统、经济系统等领域都得到了有效的应用。耗散结构理论是横跨整个自然科学和社会科学的理论,是一种普适的科学理论,具有重要的科学意义。例如在社会领域,各种团体、组织和机构都可以被看成是不同层次的耗散结构体系,都可以运用耗散结构理论来加以研究。为了维持体系的有序状态,需要提供良好的开放条件,加强与外界的物质、能量的交换以提高体系的有序度。同时需要保持体系的不平衡态来提供系统不断发展的动力,争取实现整体大于部分的非线性作用,从而促进整个社会的稳定和进化,形成高度有效的自组织结构。

**3. 耗散结构理论与体系演化**

无论是自然系统还是社会系统都会随着时间的推移不断地演化。耗散结构理论阐明:对于一个开放系统,如果它处于远离平衡态的情况下,通过系统内部诸要素的相互作用,系统自发地从无序走向有序。这个过程就是系统演化的自组织过程。系统的演化是自组织的,而不是外界和人为因素干扰的结果。自组织是事物发展过程中所选择的最优演化方式。

在自然科学领域存在自组织过程是很容易理解的，但是在社会科学领域，由于社会系统极其复杂，加之融入了人的主观情感，因此对社会科学领域的自组织过程的理解是十分困难的。而事实上，社会系统也是一个开放系统，也是耗散结构，只是在自组织程度上存在差异。封建社会时期，人的思想和行为受到压制，其自组织程度较低，社会就不会得到很好的发展，社会系统趋于混乱，最终走向衰亡；而在当今的民主法治、思想解放的现代社会里人权得到保障，人的积极性和创造性得到充分的发挥，社会能够不断地向前发展，国家就会不断走向富强。所以从耗散结构理论来看社会系统的发展，就是要不断完善耗散结构的基本条件，使社会自组织发展，逐渐从无序走向有序，进而走向更高级的有序。

**4. 耗散结构理论的方法论意义**

根据耗散结构理论的内容，我们知道，耗散结构是无法创造出来的，它的出现是无预定的，是自组织的。但是我们可以为耗散结构的出现提供有利的条件，而一旦出现了必备的条件，耗散结构的出现也是在预期之中的事了。因此耗散结构理论是系统自组织条件的理论。耗散结构理论对自然科学和社会科学都具有重要的方法论意义。耗散结构理论的运用是需要借助数学分析的方法的，但是在社会科学领域，却很少能使用数学分析方法。

由于耗散结构理论是自组织条件的理论，因此对系统自组织的研究就转化为对系统自组织条件的研究。社会系统已经是一个耗散结构，因此所要解决的问题不是耗散结构的形成而是耗散结构的完善。对于社会系统来说，我们不需要考虑系统的涨落，只需要考虑开放系统、远离平衡态和系统内各要素的非线性相互作用这三个条件就可以了。在社会系统中，开放系统主要是指信息通畅、思想开放和流通通畅等；非平衡性条件主要是指规章、制度等；非线性相互作用是指系统内各要素的相互独立、相互协调和制约等。这样我们就可以建立一个分析社会系统的思维模式，进而对社会系统进行研究。

### 12.1.6　耗散结构理论的科学价值

普利高津的耗散结构理论使人们改变了传统的看法，使其认识到自然界的过程具有随机性、不可逆性和方向性；非生命系统和生命系统一样具有从混乱向有序演化的过程。

**1. 促进各学科的发展**

在物理学领域，耗散结构理论丰富了热力学与统计物理学的相关领域的研究，为研究非平衡非线性系统问题提供了新的思路和方法。物理学家认为确定性和可逆性是物质运动的基本属性，因此就其本质而言，物质的存在是没有方向和过去的。在数学方面，耗散结构理论所描述的都是非线性的方程。耗散结构实质上就是系统在远离平衡态的一个分叉解，因此，这必然会促进分叉理论的发展与应用，这丰富了数学领域的研究内容。

在生态学方面，生态学是生态研究的一个重要内容。以往的研究主要集中在系统内各要素的平衡与相互关系上。随着耗散结构理论的引入，研究内容涉及系统内不同层次之间的结构与功能。耗散结构理论在生态学上认为，生态系统的自我调节能力依赖于系统的有序性，而有序性又以自我调节为目标，二者相互联系，共同促进系统的生态平衡与良性循环。

**2. 产生新的边缘学科**

随着耗散结构理论的发展，产生了复杂性学科、生态物理学、生物物理学等一系列边缘学科。复杂性学科主要是研究复杂性和复杂系统。生态物理学则是用物理和数学的理论去研究生态系统问题。生物物理学则是致力于研究物理学与自然界的生物之间的关系，揭示生命之谜。在各行各业中耗散结构理论都有着诱人的应用前景，随着对耗散结构理论的深入研究与应用，耗散结构理论必将为科学的发展乃至人类社会的进步做出巨大的贡献。

## 12.2 混沌现象的物理原理

世界是复杂的，它不简单，无规则，难以预测。从破碎的浪花到自由的粒子，从多姿的云朵到百态的经济，凡此种种，都是世界的复杂呈现。即便我们生活在这个复杂的世界里，但我们的目光却长久地停驻在简单的线性世界里，我们所研究的或是能研究的都逃离了“复杂”。是混沌的问世敲醒了“简单”的我们，它的兴起标志着一场探索复杂性的革命开始了。它带领我们进入复杂的世界，却又在复杂中指引我们找到出乎意料的规律性，其中分形理论为我们提供了一种发现秩序和结构的新方法。混沌与分形不仅标志着人类历史上又一次重大的科学进步，同时也正在逐渐改变人们观察和认识客观世界的思维方式。因此，对混沌与分形的深入探讨和研究具有深刻意义。

1963 年，美国著名的气象学家洛伦兹在数值实验中发现，在确定性系统中可能会表现出随机行为这一现象，他称为“决定论非周期流”。在这一论点的支配下，洛伦兹提出：“气候从本质上是不可预测的。”这个论点一直困扰着动力气象学界。后来人们发现，当时洛伦兹所发现的“决定论非周期流”现象其实就是一种混沌现象。物理学家福特(Ford)认为混沌就是 20 世纪物理学第三次最大的革命。日本著名统计物理学家久保曾说：“在非平衡非线性的研究中，混沌问题揭示了新的一页。”混沌科学最热心的倡导者之一、美国海军部官员施莱辛格这样说：“20 世纪科学将永远铭记的只有三件事：相对论、量子力学与混沌。”可见，混沌的深入研究对于人类科学的发展尤为重要。

**1. 混沌的定义**

李天岩(Li. T. Y)和约克(Yorke. J. A)于 1975 年首次提出“混沌”这一概念，并在文章“周期 3 意味着混沌”中给出了混沌的一种数学定义，现称为 Li-Yorke 定义：

设连续自映射 $f: I\to I\subset \mathbf{R}$，$I$ 是 $\mathbf{R}$ 中一个子区间。如果存在不可数集合 $S\subset I$ 满足

(1) $S$ 不包含周期点；

(2) 任给 $X_1, X_2\in S(X_1\neq X_2)$，有

$$\limsup_{t\to\infty} |f^t(X_1)-f^t(X_2)|>0$$

$$\liminf_{t\to\infty} |f^t(X_1)-f^t(X_2)|=0$$

这里 $f^t(°)=f(f(\cdots f(°)))$表示 $t$ 重函数关系；

(3) 任给 $X_1\in S$ 及 $f$ 任意周期点 $I$

$$有\limsup_{t\to\infty} |f^t(X_1)-f^t(P)|>0$$

则称 $f$ 在 $S$ 上是混沌的。

根据 Li-Yorke 定义，1983 年 Day 认为一个混沌系统应具有如下三种性质：

其一，存在所有阶的周期轨道；

其二，存在一个不可数集合，该集合只含有混沌轨道，且任意两个轨道既不趋向远离也不趋向接近，而是两种状态交替出现，同时任一轨道不趋于任一周期轨道，即该集合不存在渐近周期轨道；

其三，混沌轨道具有高度的不稳定性。

1989 年 Devaney R L 给出了混沌的另一种定义：

设 $X$ 是一个度量空间。一个连续映射 $f: X\to X$ 成为 $X$ 上的混沌，如果

（1）$f$ 是拓扑传递的；

（2）$f$ 的周期点在 $X$ 中稠密；

（3）$f$ 具有对初始条件的敏感依赖性。

除了以上提到的三种对混沌的定义外，目前还有许多其他定义，诸如 Smale 马蹄、横截同宿点、拓扑混沌、符号动力系统等定义。但至今为止，对于混沌的定义还没有一个公认的普遍适用的数学定义。

**2. 混沌的特征**

混沌是近代非线性动力学中重要的组成部分，虽然混沌的定义繁多复杂，甚至目前还尚无一个确切的定义，但混沌还是具有与其他非线性系统所不同的基本特征，具体表现如下。

（1）确定系统的内在随机性

以洛伦兹动力学方程为例，它是确定论方程，方程中不含任何随机项，方程的系数、初始条件等都是确定的，但是确定的原因却引出了随机的结果，它是由完全确定性方程产生的，一旦系统的初始状态给定，混沌信号序列便可以精确地再生，这便是混沌的特征所在。混沌系统的这一特征，洛伦兹在《混沌的本质》中说道："我用混沌这个术语来泛指这样的过程——它们看起来是随机发生的，而实际上其行为却由精确的法则决定。"洛伦兹认为，混沌系统之所以通过确定的原因产生随机的结果，这是因为"在某些动力系统中，两个几乎一致的状态经过充分长时间后变得毫不一致，恰如从长序列中随机选取的两个状态那样"。也就是说，混沌本质上是动力系统的确定性行为方式，但由于混沌运动对初值的内在敏感性存在于整个混沌区与全过程，客观上使得混沌长期行为变动不居，使得确定性混沌行为的长期预测成为不可能。所以，我们认为，混沌理论破除了"所有决定论系统都是可预测的"，否定了拉普拉斯的机械决定论。

混沌系统产生的这种类似无规则的非周期行为，可看作是确定系统的内在随机性。所谓确定系统指的是我们考虑的物理系统，它的物理量随时间的变化是一个确定性的常微分方程组或差分方程组所决定的。就处于混沌状态的动力系统来看，无论就其真实的轨道而言，还是就其相空间轨迹（如奇怪吸引子）而言，都是依据精确的规则生成的，动力系统的未来（或过去）的状态完全是由初始条件、边界条件及精确的变化规则所决定的。因而，混沌从原则上讲是完全可确定的。内在随机性，又称内禀随机性，它有别于由系统外部引入的不确定的随机影响（如噪声等）而产生的外在随机性。外在随机性是指在概率论里出现的类随机性，它是量子涨落和统计涨落的结果，这样的随机性只是表面的，其隐藏着确定论。而混沌系统中发现的这种内在随机性，与外在随机性在本质上有所不同。它是在完全确定性中引出的随机性的结果，即随机性完全是系统自身的属性，与外在因素无关，所以称之为"内在随机性"。"确定性"是因为它有内在的原因而不是外来的干扰，而"随机性"指的是不规则的不能预测的行为。所以"确定系统的内在随机性"准确地描述了混沌系统的这一特征。

（2）初值依赖性

洛伦兹在用计算机求解仿真地球大气的 13 个方程式时，为了更细致地考察结果，洛伦兹对初始输入数据的小数点后第四位进行了四舍五入，并把一个中间解 0.506 取出，提高精度到 0.506127 再送回。

当时洛伦兹的计算机使用的是纯粹决定性的方程组，经过计算后竟大吃一惊，本来很小的差异，前后计算结果却相差甚远，前后两条结果曲线几乎不存在相似性。在排除计算机的

故障问题后,洛伦兹认定,他发现了新的现象:“对初始值的极端不稳定性”,即“混沌”,又称“蝴蝶效应”,相当于亚洲的蝴蝶拍动一下翅膀,使得大气的状态产生了微小的改变,但这微小的改变将可能使美洲几个月后出现比狂风还厉害的龙卷风,又或是避免了本来会横扫印度尼西亚海岸的一场龙卷风,这就形象地说明了混沌的初值敏感依赖性。

洛伦兹关于动力系统对初始条件的敏感依赖性做出了明确的说明:第一,对初始条件具有敏感的依赖性,这并不是混沌行为的特性,其他非混沌行为同样可以具有这种属性,例如:处于不稳定平衡态的动力系统就具有对初始条件极度敏感的依赖性。一只削尖的铅笔的直立状态,就是最典型的例子。第二,“对初始条件的敏感的依赖性并不只意味着两个状态的差异随时间简单地增大”。他认为,这是由于混沌行为发生在远离平衡态的非线性区,混沌理论中所说的“‘初始条件’并非系统被构造之初存在的条件,它们经常是一个试验或计算开始时的条件,也可以是研究者兴趣所在的任一时段开始时的条件”。因此,对处于混沌行为状态的动力系统而言,对初始条件的敏感依赖性发生在整个有关的相空间中,而不是仅仅针对某一孤立的和特殊的初始状态才有意义。而且,对初始小差异的放大,不是均匀的、对称的,而是非线性的。

现在我们以一个在有限环境中世代交替的昆虫生息繁衍模型来观察混沌系统中的这种对初始值的敏感依赖性。若昆虫无外加条件控制,每年增加 $\lambda$ 倍,$N_0$ 表示环境允许的最大昆虫数量,$N_i$ 表示第 $i$ 代昆虫的总数,则有:$N_{i+1}=\lambda N_i=\lambda^{i+1}N_0$,并令 $x_i=\dfrac{N_i}{N_0}$,则用非线性迭代方程可表示为 $x_{i+1}=\lambda x_i(1-x_i)$,其中 $x_i\in[0,1]$,$\lambda\in[0,4]$。这就是有名的逻辑斯蒂映射。我们设 $\lambda$ 取定值 $\lambda=4$,给定不同的初值 $x_0$,进行迭代运算,运算数据如表 12-2-1 所示。从表中可以看出,差别如此之小的 3 个初值造成的结果差异较小,仅在小数点后 7、8 位上稍有差异,可以说前几次迭代的结果无明显差异。但经 50 次迭代后结果出现了波动,例如 50 次、51 次、52 次迭代后的结果差别甚大,似有随机性,可谓“差之毫厘,谬之千里”。

**表 12-2-1 逻辑斯蒂映射迭代运算结果**

| $N$ | $x_{i+1}=4x_i(1-x_i)$ | | |
|---|---|---|---|
| 0 | 0.1 | 0.10000001 | 0.1000001 |
| 1 | 0.36 | 0.3600000032 | 0.3600000320 |
| 2 | 0.92 | 0.9216000358 | 0.9216003584 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 10 | 0.1478365599 | 0.1478244449 | 0.1477154281 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 50 | 0.2775693810 | 0.4350573997 | 0.9372495882 |
| 51 | 0.8020943862 | 0.9831298346 | 0.1451393091 |
| 52 | 0.6349559274 | 0.0663422515 | 0.3731772536 |

(3) 奇怪吸引子

非平衡过程产生的混沌是一种“奇怪吸引子”。这是 1971 年,由法国物理学家茹勒和泰肯首次提出的另一混沌的性质。当演化时间 $t\to\infty$ 时,系统所达到的极限集合称为“吸引子”。如单摆运动,若没有摩擦或其他能量消耗,单摆将周而复始地来回摆动,运动永不停止。若单摆运动过程中有摩擦影响,即出现耗散系统,则这个状态就叫作一个吸引子。耗散系统最终要收缩到相空间的有限区域,即吸引子上。零维的吸引子是一个不动点,一维的吸

引子是一个极限环,二维的吸引子则是一个面……这些吸引子常被称为普通吸引子或平凡吸引子。混沌系统也是非平衡非线性系统演化的一种归宿,它相当于一个吸引子,它是扩散运动收缩到相空间有限区域的一种形式。但混沌系统与普通吸引子相比,又具有一些特殊的性质。系统在吸引子外的所有状态通过吸引作用都向吸引子靠拢,这是系统运动稳定的一面;但当其到达吸引子内后,其运动又是互相排斥的,这是系统运动不稳定的一面。在混沌区内,两个靠得很近的点,随着时间的推移渐渐发散开来,两个相距很远的点,慢慢地接近,它们将在混沌区中自由地游荡,却又不脱离混沌区,因此无法描写其轨迹,无法预测其未来的状态。

(4) 无周期的有序性

混沌系统具有一种无周期的有序性,并且在由分岔导致混沌的过程中,遵从 Feigenbaum 常数系。我们通过对逻辑斯蒂映射,即 $x_{i+1}=\lambda x_i(1-x_i)$,进行迭代计算,通过改变控制参数 $\lambda$ 的值来观察终态的量和质,以及终态所包含的定态的个数和大小,并判断终态是否稳定。计算表明,第一次分岔开始发生在 $\lambda=3$ 的地方,其后发生一个无穷系列的倍周期分岔,每次开始分岔的参数值为 3.449,3.544,3.564,…,其间隔越来越小,到了极限值 $\lambda=3.569$ 的地方开始进入混沌区。在 $\lambda$ 从 3.569 到 4 的参数范围内,情况极其复杂,是混沌区。但在其中有无穷多个稳定的周期解的“窗口”,窗口里又有无穷多个倍周期分岔系列,这便是通过倍周期分岔进入混沌的途径。1978 年物理学家费根鲍姆(Feigenbaum)发现了有关倍周期分岔系列的一些性质。若用 $\lambda_m$ 代表第 $m$ 次分岔出现时的 $\lambda$ 值,则相邻分岔的间距之比趋于一个常数,即有 $\lim\limits_{m\to\infty}\dfrac{\lambda_m-\lambda_{m-1}}{\lambda_{m+1}-\lambda_m}=\lim\limits_{m\to\infty}\dfrac{\Delta_m}{\Delta_{m+1}}=\delta=4.669$。此式表明,随着 $\lambda$ 的增加,两相邻分岔的间距 $\Delta_1,\Delta_2,\cdots$越来越小,倍周期的来临越来越快,这一几何级数是收敛的,收敛的比率是 4.6692,越到后面越精确。纵轴方向的分岔宽度 $\varepsilon_1,\varepsilon_2,\cdots$渐进地按照因子 $\alpha=2.5029$ 衰减,即 $\alpha=\lim\limits_{m\to\infty}\dfrac{\varepsilon_m}{\varepsilon_{m+1}}=2.5029$。也就是说,前一次的分岔宽度大约是下一次的分岔宽度的 2.5029 倍,越到后面越精确。整个系统的运行在越来越小的尺度上重复出现近似的自相似结构,由大到小的自相似的缩小比率就是一个普适的费根鲍姆数。迭代系统 $x_{i+1}=\lambda x_i(1-x_i)$的倍周期分岔现象中,分岔间距比值趋向 $\delta=4.6692$ 的情形如表 12-2-2 所示。

**表 12-2-2　周期倍增分岔情形**

| | 分岔情况 | 分岔值 | 间距比值 |
|---|---|---|---|
| 1 | 1 分为 2 | 3 | — |
| 2 | 2 分为 4 | 3.449489743 | 4.751466 |
| 3 | 4 分为 8 | 3.544091359 | 4.656251 |
| 4 | 8 分为 16 | 3.564407266 | 4.668242 |
| 5 | 16 分为 32 | 3.568759420 | 4.66874 |
| 6 | 32 分为 64 | 3.569691610 | 4.6691 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| ∞ | 周期性→混沌 | 3.569945972 | 4.669201609 |

费根鲍姆数反映出了通向混沌道路中的有序性,并揭示出倍周期分岔时分岔间距和分岔宽度分别按因子 $\delta$ 和 $\alpha$ 衰减的规律。根据费根鲍姆数可以预测参数值为多少时发生

2,4,8,…的倍周期分岔。随着控制参数 $\lambda$ 的增大,当 $\lambda$ 达到 $\lambda_\infty$ 时,周期倍增分岔就会导致"稳定的"$2^\infty$ 周期解。而 $2^\infty=\infty$,周期无穷大就意味着没有周期,因而系统从此便进入了混沌状态。正是因为混沌是有规律的,非混乱的,所以有这种说法:"混沌是非周期的有序性",又或是"混沌是确定性的非周期流"。也就是说,混沌并非简单的有序态,而是一种没有确定周期性和明显对称性的有序态。

(5) 有界

一般我们所研究的或能研究的物理问题都是本质上的回归行为,其现实的测度空间都有确定的边界,那些无界的本质上的非回归行为则没有一般意义。所以,我们所面对的现实系统的状态变量不可能是绝对的无穷大,都是局限于确定范围内的有限值,即"有界"。并且,混沌系统的局部不稳定性并不意味着整体是不稳定的,实际上混沌也有稳定性的一面,它的整体稳定性是混沌系统的一个重要性质,是混沌系统所具有的稳定机制的反映。

我们以抛物线映射为例,实际上它就相当于特定的拉伸与折叠变换,拉伸操作改变了系统的初始条件,导致了混沌的不稳定性,而折叠操作却使得系统的取值空间减小了一半,把相点始终限制在最初的映射区间内,这使得混沌具有一定的稳定性。所以,本质上的回归行为的现实物理系统的稳定机制及表观上的整体稳定性就是物理上的"有界",也就是混沌的一个本质特征。

例如,在 Li-Yorke 定义中,"有界"是定义的前提条件,它设定了 $f$ 是从 $I$ 到 $I(IR)$ 的映射,而 $I$ 是 **R** 中的一个闭区间,这表示 $f$ 把 $I$ 映射回 $I$,所有的相点无法超越 $I$ 的确定边界,这个"有界"的前提条件的设定是必要的,如若没有这个前提条件,就无法保证系统是混沌的,例如,映射 $f$: $X_{i+1}=f(X_i)=X_i^2$,当 $X_1>1$ 时,$f$ 会很快使 $X_i$ 超越 $I$ 的边界并趋于 $\infty$,虽然它仍具有非周期性、对初始条件的敏感依赖性等混沌的本质特征,但是它的演化过程是发散的,不能形成混沌吸引子。所以,"有界"是形成混沌的必要条件和本质特征。

## 12.3 分形理论简介

分形理论是非线性科学的前沿和重要分支,它是近三十多年来科学前沿领域提出的一个极为重要的概念,人们将其与耗散结构、混沌并称为 20 世纪 70 年代科学史上的三大发现。作为一种方法论和认识论,其启示是多方面的:一是分形整体与局部形态的相似,这启发人们通过认识部分来认识整体,从有限中认识无限;二是分形揭示了介于整体与部分、有序与无序、复杂与简单之间的新形态和新秩序;三是分形从一特定层面揭示了世界普遍联系和统一的图景。

### 12.3.1 分形理论概述

分形理论是一门交叉性的横断学科,从振动力学到流体力学、天文学和计算机图形学,从分子生物学到生理学、生物形态学,从经济学到语言学、社会学,从材料科学到地球科学、地理科学等,分形理论无处不在。从分形的角度看世界,我们发现,这个世界其实是以分形的方式存在和演化着的。

分形理论诞生于美国 IBM 的研究人员曼德布罗特(B. B. Mandelbrot)1967 年在"*Science*"杂志上发表的一篇题为《英国的海岸线有多长?统计自相似性与分数维数》(*How Long is the Coast of Britain*,*Statistical Self Similarity and Fractional Dimension*)的论文中。

曼德布罗特在其两本著作《自然界的分形几何学》和《分形：形状、机遇与维数》中首次提出了“fractal”一词，其原意是“不规则的”“分数的”“支离破碎的”物体，并阐述了分形理论的基本思想，即分形研究的对象是具有自相似性的无序系统，其维数的变化是连续的，因而又可称为“碎形”。

对于分形理论的非线性现象十分普遍又难于研究。说到普遍，也许大多数人不知道如此普遍的分形究竟是什么。其实，西兰花、云朵、树木、山川、粒子的布朗运动、脑电图等都是分形的具体呈现，或许你花上几分钟仔细地观察一下西兰花就会大致明白什么是分形了。但分形理论却不像其表观那样容易观察和研究，其实对于分形，至今尚缺少一个确切的定义。

曼德布罗特曾经为分形下过两个定义。

**定义 1**：若一个集合在欧氏空间中的 Hausdoff 维数 $D_h$ 恒大于其拓扑维数 $D_t$，即 $D_h > D_t$，则称该集合为分形集，简称分形。

**定义 2**：部分与整体以某种形式相似的形，称为分形。

定义 2 突出了分形的自相似性，反映了自然界中广泛存在的一类事物的基本属性：局部与局部，局部与整体在形态、功能、信息、时间、空间等方面具有统计意义上的自相似性。它与欧氏几何中的“相似”不同。

### 12.3.2　分形理论的特征

经过理论的验证和实际的检验，人们发现以上两个定义难以包括分形如此丰富的内容，它们仍然不是最严密、最精确的定义。人们发现正如生物学中对“生命”也没有严格明确的定义一样，我们通常是列出生命体的一系列特征来加以说明。对分形的定义，我们也可以这样处理。通过对分形的观察与深入研究，我们发现分形具有以下五种特征。

**1. 形态的不规则性**

无论从局部看还是从整体看，分形都不是规则的，它既不是满足某些简单条件的点的轨迹，也不是任何简单方程的解集，所以分形不能通过传统的几何语言进行描述。当然，分形也无法通过简单的长度、面积、体积等测量尺度来度量。例如，由于涨潮落潮使海岸线的水陆分界线具有各种层次的不规则性，海岸有平坦的沙滩，但更多的是岩石块、结构各异的海湾、大小不一的断层、深浅不同的峡谷，其结构十分不规则，即使是精度再高的地图也无法将其细节一一详细展露。海岸线在大小两个方向都有自然的限制，取不列颠岛外缘上几个突出的点，用直线把它们连起来，得到海岸线长度的一种下界，使用比这更大的尺度是没有意义的。还有海沙石的最小尺度是原子和分子，使用更小的尺度也是毫无意义的。在这两个自然限度之间，存在着可以变化许多个数量级的“无标度”区，既然长度不是海岸线的定量特征，就要用分维。也就是说，分形具有许多个标度和无标度区，分形的长度不能精确地测量出来，更无法用简单的几何语言来描述。

**2. 结构的精细性**

分形具有的普遍特征是其结构具有无限的精细性。即具有任意小比例的细节，在分形上任选一局部区域，无论将其放大或缩小，其形态、复杂程度、不规则性等各种特性均不会发生变化，故又可称为标度不变性或是伸缩对称性，即这类物体没有特征长度。特征长度是指所考虑的对象中最具代表性的尺度，比如时间的分、秒、时，及空间的长、宽、高等。具有自相似性的系统、物体、事物必定满足标度不变性，或者说这类形体没有特征长度——没长短、面积、体积等。对于实际的分形体来说，这种标度不变性只在一定的范围内适用。我们通常把

标度不变性适用的空间称为该分形体的无标度空间,其内是分形,在此范围以外就不是分形了,它有有限与无限之分。

**3. 局部与整体的自相似性**

局部与整体的自相似性是许多分形的重要特征,它指的是复杂系统的总体与部分,这部分与那部分之间的精细结构或性质所具有的相似性,或者说从整体中取出的局部能够体现整体的基本特征,即几何或非线性变换下的不变性,也就是在不同放大倍数上的形状相似。

自相似性的数学表示为 $f(\lambda r)=\lambda^{\alpha} f(r)$。其中 $\lambda$ 为标度因子,$\alpha$ 为标度指数,它描述了结构的空间性质。函数 $f(r)$是面积、体积、质量等占有数、量等性质的测度。

一般情况下自相似性有比较复杂的表现形式,并不是局域放大一定倍数后与整体完全重合的简单行为。但表征自相似系统或结构的定量性质如分形维数,并不会因放大或缩小等操作而改变,所改变的只是其外部的表现形式。

自相似性是从不同尺度的对称出发,也就意味着递归。分形体中的自相似性可以是完全相同,也可以是统计意义上的相似。标准的自相似分形是数学上的抽象,迭代生成无限精细的结构,这种有规则的分形只是少数,绝大部分分形是统计意义上的无规则分形。简单地说,就是若一个图形的每个局部都可以放大为整个图形,则称整个图形具有自相似性。

Cantor 集是典型的自相似分形,如图 12-3-1 所示;Koch 雪花也具有自相似性,如图 12-3-2 所示;当然还有 Sierpinski 地毯曲线,如图 12-3-3 所示。甚至自然数的 Sharkovsky 序也具有自相似性,该序结构的每个片断理解为每一点有规律的后继,显然每一片断在除以 $2^k$ 因子,即完成一相似变换后又得到完整的 Sharkovsky 序,除非这个片断是有限的。自然界很多事物都是具有自相似性的,例如叶脉、树枝、根系等。

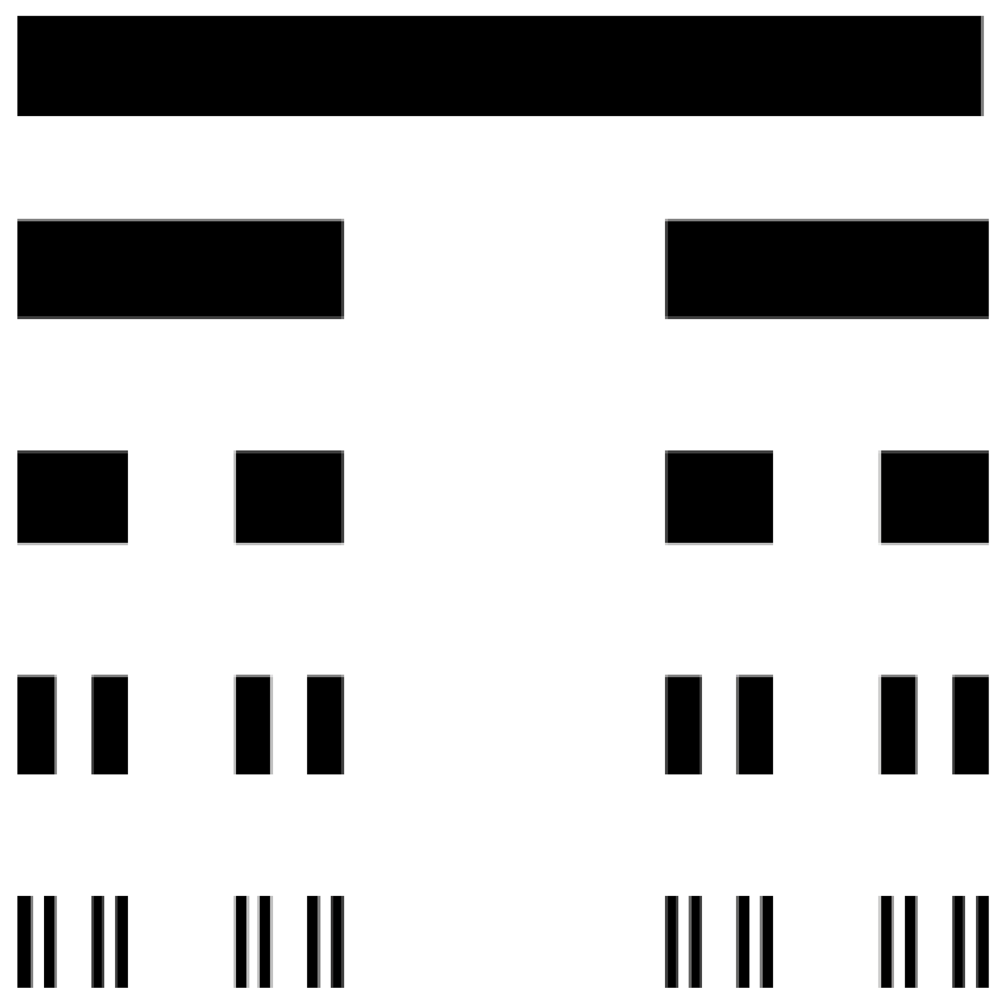


图 12-3-1 Cantor 集

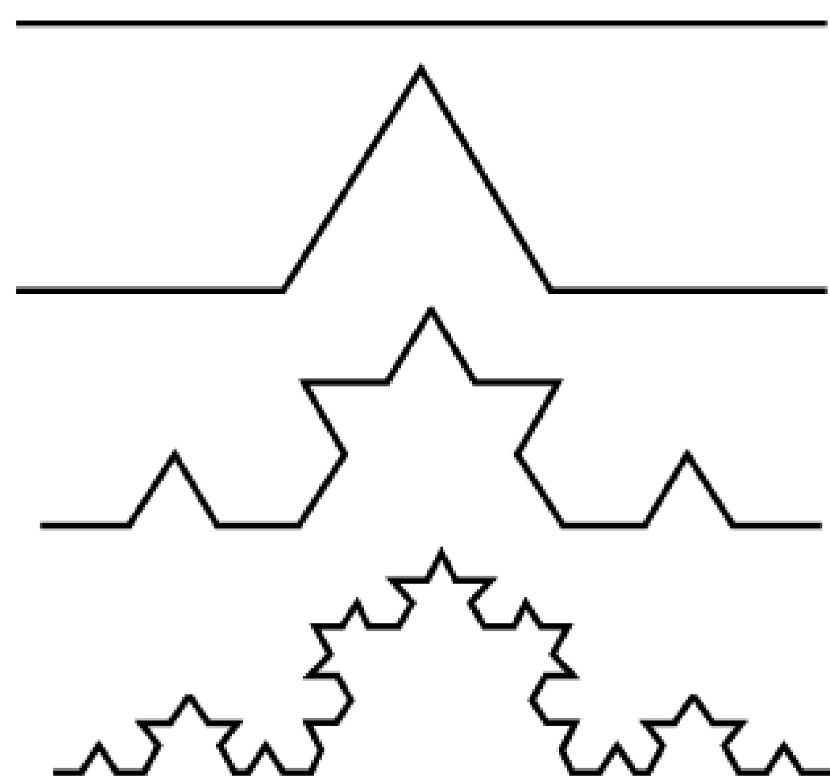


图 12-3-2 Koch 雪花

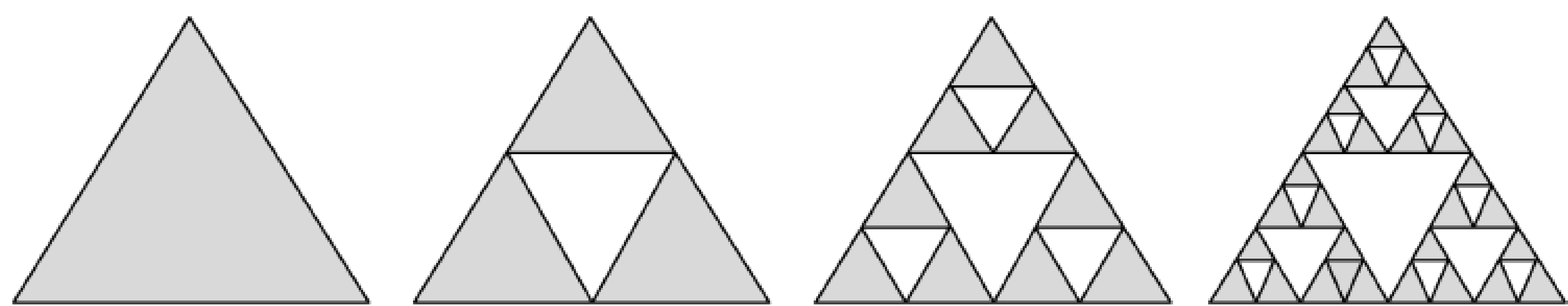


图 12-3-3 Sierpinski 地毯曲线

**4. 维数的非整数性**

非整数的 Hausdorff 维数是分形的一个重要特征。维数表示描述一个对象所需的独立变量的个数。我们习惯于将点定义为零维,直线为一维,平面为二维,空间为三维。爱因斯

坦在相对论中引入时间维，便形成了四维时空。我们对某一问题给予多方面的考虑，即可建立高维空间，但都是整数维。数学上将欧氏空间的几何对象连续地拉伸、压缩、扭曲，维数也不变，这就是拓扑维数。然而，曼德布罗特曾描述过一个绳球的维数：从很远的距离观察这个绳球，可看作一点，即零维；从较近的距离观察，它充满了一个球形空间，即三维；再近一些，就看到了绳子即一维；再向微观深入，绳子又变成了三维的柱，三维的柱又可分解成一维的纤维，再近，则又变为三维的柱了。所以说对象的维数是可以变化的，关键是我们从什么尺度去观察它，一旦确定了尺度，那么对象的维数也就确定了。其实，数学家豪斯道夫(Hausdoff)早在1919年就提出了连续空间的概念，也就是空间维数是可以连续变化的，它可以是整数也可以是分数，这称为Hausdoff维数，记作$D_f$，一般的表达式为：$K=LD_f$，取对数并整理得$D_f=\dfrac{\ln K}{\ln L}$，其中$L$为某客体沿其每个独立方向皆扩大的倍数，$K$为得到的新客体是原客体的倍数。显然，$D_f$在一般情况下是一个分数。所以，曼德布罗特也把分形定义为Hausdoff维数大于或等于拓扑维数的集合。我们之所以根据欧氏一维测度无法计算海岸线长度正是由于欧氏一维测度与海岸线的维数不一致，而利用Hausdoff维数可得知英国的海岸线维数为1.26，进而海岸线的长度也就确定了。

**5. 生成的迭代性**

在实际应用上，要生成分形可以依靠图形或函数的递归来实现。递归是指一个过程直接或间接地调用其自身的一种行为。首先给定初始图形$F_0$，依照某一规则$R$对图形反复作用，即可得到

$$F_{k+1}=RF_k,\quad k=0,1,2,\cdots$$

进而得到图形序列$F_1,F_2,\cdots$，其极限图形就是分形，作用规则$R$成为分形元，如图12-3-4、图12-3-5和图12-3-6所示。

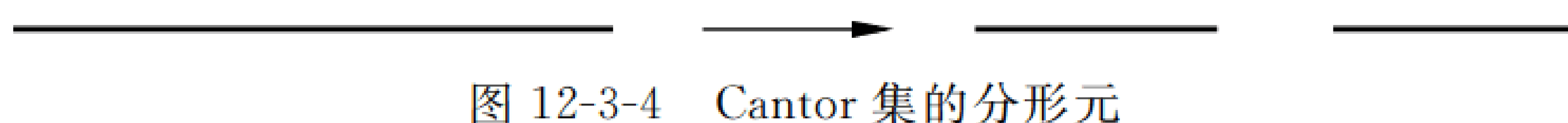


图12-3-4　Cantor集的分形元



图12-3-5　Koch雪花的分形元

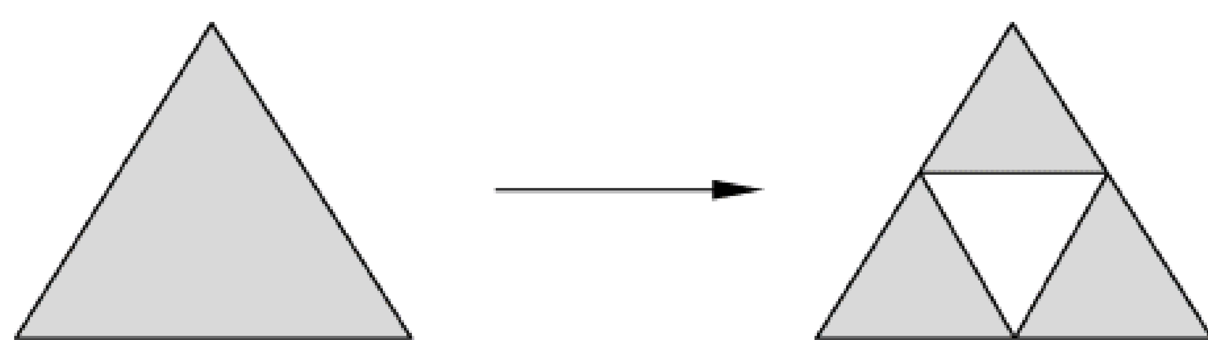


图12-3-6　Sierpinski地毯曲线的分形元

### 12.3.3　混沌与分形的联系与区别

在非线性科学中，混沌与分形有着不同的起源，混沌产生于物理学的研究，而分形则来源于几何学的研究，但它们都是非线性方程所描述的非平衡的过程和结果，这意味着它们有共同的数学祖先，即动力系统。结构的复杂性使得现实世界出现了大量分形几何形体，同时也使确定性动力学体系出现无规则性。混沌吸引子就是分形集，也可以说混沌是时间上的分形，而分形是空间上的混沌。混沌的奇怪吸引子都具有自相似性，无穷相似结构相互层叠

起来,结构便失去了规则,导致“无穷嵌套的自相似结构”呈现出总体的混沌。非线性动力学系统一旦进入混沌吸引子区域,便会随机地在吸引子内部到处游荡,却又不能充满整个混沌吸引子区域,因而区域内存在着无穷多的随机空隙,这使整个混沌区出现维数上的“空洞”,呈现出分数维数。洛伦兹吸引子就是三维背景空间中的一张分形曲面,其容量维等于2.06,若斯勒吸引子同样也是三维背景空间中的一张分形曲面。所以,分形几何学和分维在混沌学研究的舞台上扮演着重要的角色。

混沌与分形理论的联系,多是以自组织系统为其研究对象的。自组织现象,通常是时空有序的结构,是复杂的系统,无法用传统的简化方法来解决,所以,需要依靠新的研究复杂性的方法来处理,混沌与分形就首当其冲。混沌中有时含有分形,而分形中有时又孕育着混沌。但不同的是,分形更注重形态或几何特性及图形的描述,而混沌则更偏向数理的动力学及动力学与图形结合的多方位的描述和研究。分形更重视有自相似性的系统,而混沌涉及面似乎更广,对所有的有序与无序现象都感兴趣。所以,它们是从不同侧面来研究同一个问题的。

总的来说,混沌与分形的起源和发展过程不相同,但它们的研究内容从本质上讲存在着一定的相似性,混沌主要在于研究过程的行为特征,分形更关注吸引子本身结构的研究。尤其是混沌中的分岔、分支现象与分形关系最密切,但有些混沌系统自相似性未必特别显眼,那么分形恐怕就难以涉足了。总之,想要较确切和较系统地说明混沌与分形的联系与区别,目前还比较困难,这有待混沌与分形理论进一步的深入研究与拓展。

### 12.3.4 非线性科学的前景与展望

现代科学所面临的是“简单”的思维和方法无法处理“复杂”的客观问题。因而,一系列以复杂系统为研究对象的学科相继出现,它们逐步改变了人们观察和思考世界的方法,并启发人们不断突破不同科学领域的局限性,从而形成了一门崭新的综合性前沿学科,即以混沌理论、分形几何学和孤立子理论为主题的非线性科学。面对多姿多彩的客观世界,人们渐渐发现绝大部分的自然现象和数学方程其实都是非线性的,那些看似可以描述它们的线性方程其实正是因为它们太复杂了。随着人类科学的发展,非线性科学的研究已备受瞩目,无论是自然科学、工程技术,还是社会科学等各领域,关于非线性问题的研究从未停止过,以后也不会停止。

虽然非线性科学的发展并不久远,但我们仍能看到非线性科学闪耀的光芒:耗散结构和孤立子解释了非线性作用引起的惊人的有序性;确定性系统的混沌使人们看到了普遍存在于客观世界而人们一直忽视的随机现象;分形和分维的研究将人们从线、面、体的传统几何观念中解放出来……这些都是目前我们在非线性科学领域取得的成就。随着非线性科学的发展,人们可以将混沌运用于系统辨识、最优参数设计等,在神经网络的运用方面,混沌的发展有利于联想记忆和机器人的路径规划等发展,混沌还可以实现图像数据压缩、高速检索,并且我们可以通过混沌现象的决策论实现对非线性时间序列的预测……而分形理论也同样广泛运用于我们的日常生活。

将来我们可以利用分形理论来推测结团表面化学物质的吸附量来优化空气质量,我们也可以通过分形理论的发展推动生物医学的发展,分形理论更是对材料中的催化问题的研究开辟了新的道路。可见,非线性科学的发展前景是非常光明的,但非线性科学的世界枝繁叶茂,我们想要完全了解非线性世界,形成一套完整的科学体系,还需要很长时间,在非线性科学的发展道路上还有许多未知的困难与阻碍,这就要求我们要有坚定的科学探究精神和

坚实的科学文化素养，在非线性科学的道路上昂首并进，以“复杂”应“复杂”，一同展望非线性科学的蓬勃发展。

# 学习资源

## 一、中国近现代著名科学家（物理学）简介

王竹溪（1911—1983），名治淇，字竹溪，以字行，湖北省公安县人，物理学家、教育家、中国热力学统计物理研究开拓者。1933年7月毕业于清华大学物理系；1935年7月毕业于清华大学研究院，1938年7月获剑桥大学哲学博士学位，1952年加入九三学社，1955年当选为中国科学院院士，1979年参加中国共产党。



王竹溪出身于书香门第，父亲和祖父都是清末秀才。他幼年在家乡曾先后随其舅父及长兄治昆等读四书、《诗经》、《左传》等书。1922年春入麻豪口小学，1922年秋到武昌入荆南中学补习班，1923年升入正式班。1926年夏毕业。在1922到1926年期间他除完成学校的课程外，还博览群书。曾先后学习《幼学琼林》《庄子》《老子》《荀子》《礼记》《史记》《汉书》等书，对《康熙字典》的切韵法颇感兴趣。他对数学特别爱好，喜读中国名家小说。从而奠定了少年启蒙知识的基础。

王竹溪主要从事理论物理特别是热力学、统计物理学、数学物理等方面的研究。在湍流尾流理论、吸附统计理论、超点阵统计理论、热力学平衡与稳定性、多元溶液、热力学绝对温标、热力学第三定律、物质内部有辐射的热传导问题以及基本物理常数等领域取得多项重要成果。在有序无序变化的统计力学理论方面将贝特理论作了重要推广，在热力学的理论研究方面做出多方面的推广。同时对物理学史、基本物理常数和汉字检索机器化方案等作了不少有成效的研究。他还在生物学关键问题-水分化学势、汉字检索方案的优化等方面做出了突出贡献。

赵忠尧（1902—1998），中科院院士、核物理学家，中国核物理的鼻祖，中国核事业的先驱之一。浙江诸暨人，1920年考入南京大学前身南京高等师范学校。1925年毕业于南京东南大学化学系，并入清华大学任教。1927年赴美国加利福尼亚州理工学院留学，1930年获哲学博士学位，并赴德国哈罗大学进行科学研究。1931年回国后曾任清华大学物理系教授、云南大学教授、西南联合大学教授、中央大学教授及物理系主任。1946年赴美国进行核物理和宇宙线方面的研究。1949年，在美国加州理工学院进行原子核反应研究。1950年回国，带回当时国内尚无条件制备的器材和设备。1955年，赵忠尧用带回的器材和零件，主持建成了中国第一台质子静电加速器，并进行了原子核反应的研究。9年之后，中国第一团“蘑菇云”在祖国大西北升空，是第一个看到“蘑菇云”的中国科学家。



赵忠尧用带回的器材和零件,主持建成了我国第一台70万电子伏的质子静电加速器。1958年又主持研制成功250万电子伏的质子静电加速器。这两项研究的成功,对我国的核事业具有举足轻重的作用。1984年,北京正负电子对撞机工程破土动工。1989年以来,一批新的科研成果陆续问世,这一切蕴含了包括赵忠尧在内的老一辈科学家的心血,而此时,奋战在科研一线的则是他们培养出来的新一代中青年科学家。历任中国科学院近代物理研究所、原子能研究所、高能物理研究所研究员、副所长,中国科技大学教授、近代物理系主任,中央研究院院士,中国物理学会副理事长、名誉理事,中国核学会副理事长、名誉理事长等职。

## 二、诺贝尔物理学奖简介

**2001年诺贝尔物理学奖**。取得了在淡气中实现碱性原子的博斯-爱因斯坦冷凝,揭示了一种新的物质状态:博斯-爱因斯坦冷凝物。获奖者:艾里克A.科纳尔(美国)、沃尔夫冈.凯特纳(德国)以及卡尔E.威依迈(美国)。评价:人们都知道激光与普通灯光不同,在激光中所有光粒都具有相同的能量与振荡,因此如何控制其他光束达到同样的状态一直就是对物理科学家的一种挑战。

**2002年度诺贝尔物理奖**。对天体物理开创性的贡献,特别是对宇宙中微子的探测。获奖者:雷蒙德·戴维斯(美国)、小柴昌俊(日本)和里卡尔多·贾科尼(美国)。评价:在天体物理学领域做出的先驱性贡献,其中包括在“探测宇宙中微子”和“发现宇宙X射线源”方面的成就。科学家雷蒙德·戴维斯,他因为在天体物理学,特别是“探测宇宙中微子”领域做出的先驱性贡献,获得今年诺贝尔物理学奖1/4的奖金。日本科学家小柴昌俊与美国科学家雷蒙德·戴维斯相似的贡献,得以分享今年诺贝尔物理学奖1/4的奖金。美国科学家里卡尔多·贾科尼现为负责美国国家射电天文台运作的大学联合公司(AUI)主任。

**2003年诺贝尔物理学奖**。超导体和超流体方面的理论成果。获奖者:阿布里科索夫(俄罗斯)、金兹布尔格(俄罗斯)和利盖特(英国,后加入美国籍)三位科学家。评价:由于在创立解释量子论中的两大现象:超导体和超流体方面的理论方面做出突出贡献而获得这一荣誉。这三位获奖者都是高龄老人,他们将分享总额为130万美元的奖金。阿布里科索夫1928年生于莫斯科,曾在俄罗斯的多家研究所和大学工作,涉足超导体、金属、半金属和半导体等多个研究领域。金茨堡1916年生于莫斯科,他在20世纪50年代与一些科学家率先提出了有关Ⅱ型超导体的理论。莱格特1938年生于伦敦,现为美国伊利诺伊大学厄巴纳-尚佩恩分校教授。

**2004年诺贝尔物理学奖**。在夸克粒子理论方面取得了成就。获奖者:戴维·格罗斯(美国)、戴维·波利泽(美国)和弗兰克·维尔泽克(美国)三位科学家。评价:夸克是自然界中最小的基本粒子。这三位科学家对夸克的研究,使科学更接近于实现,它为“所有的事情构建理论”的梦想,至少在“最微小的世界”里,人们实现梦想的旅程前进了一大步。

**2005年诺贝尔物理学奖**。通过梳状滤波技术,所有颜色的光谱都可以被测量。获奖者:罗伊·格劳伯(美国)、约翰·霍尔(美国)和西奥多·汉施(德国)。格劳伯将获得总值1000万瑞典克朗奖金的一半。霍尔和汉施两教授将各获得奖金总额的1/4。评价:在光学领域做出巨大贡献:格劳伯奠定了量子光学的理论基础,据此,光中微粒子的运动可以被描述。而霍尔和汉施对精密光谱学的研究,使得人们可以精确地测得原子和分子的光学颜色,还可以把对光频率的测量精确到15位数。格劳伯1925年出生,现为美国哈佛大学教授。霍尔生于1934年,现为美国科罗拉多大学教授。汉施1941年出生,现为德国慕尼黑麦克

斯-普兰克量子光学研究所研究员。

**2006 年度诺贝尔物理学奖**。宇宙背景辐射的黑体形式和各向异性。获奖者：约翰·麦瑟(美国)和乔治·斯姆特(美国)。评价：发现了宇宙背景辐射的黑体形式和各向异性，宇宙微波背景辐射的黑体谱形状及其温度在不同方向上的微小变化。他们的出色工作被誉为是宇宙学研究进入精密科学时代的起点其成果有助于研究早期宇宙，并能帮助人们更多地了解星系和恒星的起源。“他们并没有证明‘大爆炸’理论，但为这一理论提供了强有力的支持。这是本世纪最伟大的发现之一，它扩展了我们对宇宙的认识。”麦瑟是美国航天局戈达德空间飞行中心的一位科学家；斯姆特今年在加州大学伯克利分校的劳伦斯伯克利国家实验室工作。

**2007 年度诺贝尔物理奖**。巨磁电阻效应。获奖者：阿尔贝·费尔(法国)和彼得·格林贝格尔(德国)，表彰他们发现巨磁电阻效应的贡献评价：该成果推动了磁信息材料的发展，$MgFe_2O_4$ 是一种磁信息材料，广泛应用于电子计算机。

**2008 年度诺贝尔物理奖**。亚原子物理学中的自发性对称破缺机制。获奖者：南部阳一郎(日本，美国籍)、小林诚(日本)和益川敏英(日本)三位科学家。评价：发现亚原子物理学中的自发性对称破缺机制。对称破缺包含“自发对称破缺”和“动力学对称破缺”两种情形。3 名物理学家分别在 20 世纪 60 年代和 70 年代通过数学模型“预言”了量子世界自发性对称破缺现象的存在机制和根源。预言三种夸克的存在。“小林-益川理论”也因此成为支撑亚原子物理学标准理论的重要支柱。这是 2002 年小柴昌俊获诺贝尔物理学奖以来，事隔 5 年日本学者再度摘得物理学奖；也是 1949 年汤川秀树以来，日本学者第 5 次获得诺贝尔物理学奖。

**2009 年诺贝尔物理奖**。光在纤维中的传输以用于光学通信成果。获奖者：高锟(中国，英国籍)以及威拉德·博伊尔(美国)和乔治·史密斯(美国)。评价：高锟在“有关光在纤维中的传输以用于光学通信方面”取得了突破性成就，他将获得今年物理学奖一半的奖金，共 500 万瑞典克朗(约合 70 万美元)；博伊尔和史密斯发明了半导体成像器件——电荷耦合器件(CCD)图像传感器，将分享今年物理学奖另一半奖金。高锟，1933 年 11 月 4 日年出生在上海金山，拥有英国和美国双重国籍的物理学家、北京邮电大学名誉教授。1966 年，高锟发表了一篇题为《光频率介质纤维表面波导》的论文，开创性地提出光导纤维在通信上应用的基本原理，描述了长程及高信息量光通信所需绝缘性纤维的结构和材料特性。由于他在光纤领域的特殊贡献，获得巴伦坦奖章、利布曼奖、光电子学奖等，被称为“光纤之父”。1987 年 10 月，高锟从英国回到香港，并出任香港中文大学第三任校长。于 1996 年当选为中国科学院外籍院士。1996 年中国科学院紫金山天文台将一颗于 1981 年 12 月 3 日发现的国际编号为“3463”的小行星命名为“高锟星”。

**2010 年诺贝尔物理学奖**。二维空间材料石墨烯(graphene)方面的开创性实验成果。获奖者：英国曼彻斯特大学 2 位科学家安德烈·盖姆(英国)和康斯坦丁·诺沃肖罗夫(英国)。评价：他们在二维空间材料石墨烯(graphene)方面的开创性实验成果。作为由碳组成的一种结构，石墨烯是一种全新的材料——不单单是其厚度达到前所未有的小，而且其强度也是非常高。同时，它也具有和铜一样的良好导电性，在导热方面，更是超越了目前已知的其他所有材料。石墨烯近乎完全透明，但其原子排列之紧密，却连具有最小气体分子结构的氦都无法穿透它。碳——地球生命的基本组成元素——再次让世人吃惊。安德烈·盖姆和康斯坦丁·诺沃肖罗夫是从一块普通得不能再普通的石墨中发现石墨烯的。他们使用普

通胶带获得了只有一个原子厚度的一小片碳。而在当时,很多人都认为如此薄的结晶材料是非常不稳定的。

## 思考与练习

1. 熵增原理与热力学第二定律的关系及其启示。
2. 耗散结构与热力学第二定律的关系及其启示。
3. 讨论混沌现象原理。
4. 讨论分形理论原理。
5. 讨论分形和混沌的关系。
6. 谈谈非线性科学的基本特征与研究展望。

# 第13章

# 大学物理学中的数学方法

德国哲学家康德(Immanuel Kant,1724—1804)说过:“任何一门科学,只有当它数学化之后,才能称得上真正的科学。”美国著名科学史教授戴维·林德伯格(D. Lindberg)在《西方科学的起源》一书指出:“有关科学的几个概念已得到有力的支持,其中之一的概念为:依据理论的陈述形式来定义,陈述的形式应当是一般的、定律式的,最好用数学语言来表述。”冯友兰在《中国哲学史》注中说:“……科学所说之事物,亦在语言文字之外;然此等事物,只是事物,不是科学;语言文字所表之原理公式等,方是科学。”

如果说只有能用语言文字表述为原理公式的,才能称之为“科学”,那么,物理学的“语言文字”就是数学,物理学理论的严密性、精确性和自洽性被完美地体现在数学语言之中。因此,物理学是一门定量精密的科学。从物理概念转变为物理量开始,物理学的基本定律与原理,是物理量之间的函数关系在一定条件下规律性反映。数学作为研究物理学的一种重要方法和工具,为物理学提供了描述物理概念和规律的简洁、精确、形式化的语言和表达式,提供了对观测材料进行科学抽象的手段,简化和加速了物理思维进程,促进了物理学理论的建立和发展。

## 13.1　力学中的极限与微积分

每一学科的深入发展都离不开数学符号的使用和数学理论的严密推演(关于数学的本质,请参阅本书第2.9节)。数学是物理学家的思维工具,只有通过数学才能最终以精确的形式表达自然规律,只有通过数学才能抓住错综复杂的变化过程,从中找到最基本、最普遍的规律,只有通过数学才能将物理学构建成严密系统的理论体系。例如:没有对数,开普勒难以建立天文学的重要规律——开普勒三定律;没有微积分,牛顿得不出万有引力定律;没有统计学,无法发展分子动理论;没有黎曼张量,爱因斯坦的广义相对论不能完善。总之,物理学的发展离不开数学,而数学的发展也和物理学密切相关。当然,物理学不等于数学,物理学的发展也不仅仅靠数学,数学只是一种形式逻辑,物理学发展还需要物理实验和生产技术的检验和推动。

### 13.1.1　力学与微积分

**1. 力学的发展与微积分的发明**

17世纪的欧洲科学技术发展迅速、成果丰富,而最辉煌的成就是微积分的发明,它的出现是整个数学史也是整个人类历史上的一件标志性事件。微积分的使用,使人类找到了认识自然界相互联系、变化和发展的最本质理论工具,开拓了人类的思维方式和视野,深刻地促进着自然科学和生产技术的发展。

在微积分的发明上,莱布尼兹(Leibnitz Gottfried Wilhelm von,1646—1716)和牛顿功

绩相当。这两位数学家在微积分学领域中的卓越贡献概括起来主要是：他们总结出处理各种有关问题的一般方法，分析了极限的本质含义，认识到求积问题与切线问题互逆的特征，并揭示出微分学与积分学之间的本质联系；他们都各自建立了微积分的基本定理，他们给出的微积分的概念、法则、公式和符号为以后微积分的进一步发展奠定了坚实的重要的基础。总之，他们创立了作为一门独立学科的微积分学。

牛顿与莱布尼兹创立微积分的理论基础存在着差异。牛顿注重物理方面，以连续运动为出发点，因而具有比较明显的极限概念；莱布尼兹侧重于几何方面，他以离散的无穷小为出发点，因而极限的观念不甚鲜明。牛顿是在研究力学问题时感到许多变量和特殊形状物体的面积(比如分析开普勒定律时的椭圆面积)或体积等的计算问题无法解决而研究微积分，所以，牛顿的微积分理论有着明显的力学背景，下面主要介绍牛顿的微积分。

牛顿微积分的起源是运动学，牛顿在《流数简论》中借助于运动学中描述的连续量及其变化率阐述他的流数理论。1666年，他在坐标系中通过速度分量来研究切线，促使了流数的产生，并提供了它的几何应用的关键。牛顿把曲线 $f(x,y)=0$ 看作沿 $x$ 轴与 $y$ 轴运动的点的轨迹，动点的坐标 $x,y$ 是时间的函数，$\dot{x}$ 表示动点的水平速度分量，$\dot{y}$ 表示垂直速度的分量。他把 $\dot{x}$ 与 $\dot{y}$ 随时间变化的“流动速度”称作“流数”。实际上就是 $x$ 和 $y$ 对 $t$ 导数 $\dot{x}=\frac{\mathrm{d}x}{\mathrm{d}t},\dot{y}=\frac{\mathrm{d}y}{\mathrm{d}t}$。但当时牛顿并没有用这些符号表示，牛顿创立了用字母上加一点的符号的表示法也是晚些时候才引入的。他提出了两大问题：已知 $x,y$ 之间的关系 $f(x,y)=0$，求流数 $\dot{x},\dot{y}$ 之间的关系。反过来，已知 $x$ 与 $y/x$(切线斜率)之间的关系，求 $x,y$ 之间的关系。牛顿提出的方法与以前的方法是根本不同的，他首先确定所求面积对横坐标的变化率，再通过“反微分”求出面积。这种做法实质上就是将面积计算看作是求切线的逆过程，从而确立了这两种运算的互逆关系。牛顿把这种互逆关系作为一般规律明确地揭示出来，完成了微积分的最后也是最关键的一步。微分和积分过程的互逆性今天称为微积分的基本关系。

在科学史上，还没有一个人能像牛顿那样以自己的数学发现而对近代科学产生如此广泛而深刻的影响。就数学本身而言，牛顿的发现(当然与莱布尼兹一起)开辟了全新的天地。在这以前，数学就是几何与代数，微积分的进入则使这门科学在内容、意义和方法上都发生了根本的变革。伴随着微积分的进一步发展而产生的微分几何、微分方程、变分法等一系列新的分支(可以说牛顿也同时是几乎所有这些分支的先驱者)，与微积分本身一起形成了数学中被称之为“分析”的广大领域，并与古老的几何与代数三足鼎立，并且在18世纪，其繁荣程度就远远地超过了几何与代数。

**2. 力学和微积分**

牛顿的数学成就——微积分，开拓了崭新的近代数学，他还依靠他所创立的数学方法实现了自然科学的一次巨大综合，并奠定了我们今天所说的经典力学体系的基础。

牛顿被誉为科学的巨人。纵观牛顿的一生，他在科学上最重要的成就有三个：发明微积分、建立经典力学体系、提出光的性质理论。其中任何一项成就都足以使他列入大科学家的行列。1687年，牛顿撰写的《自然哲学的数学原理》出版。这部书是人类的自然科学的奠基性巨著，是自然科学史上最重要的著作之一。著名的牛顿运动三定律、万有引力及牛顿的微积分成果都载于此书。

爱因斯坦曾指出：“……命运使他(牛顿)处在人类理智的历史转折点上。”牛顿的数学

成果,也正是17世纪伟大时代的产物。

为了理解这一背景,不妨略微回顾一下牛顿1642年诞生之前大约半个世纪里自然科学的形势。在这短短不到50年的时间里,天文、力学和光学三大领域重大事件接踵而至。世纪之初,望远镜的发明不仅引起了天文学的新高涨,而且推动了光学的研究;1619年开普勒公布了他的最后一条行星运动定律,从数学上推证开普勒基于观测获得的经验定律,成为当时自然科学的中心课题之一;1638年伽利略《关于两门新科学的对话》的正式出版,激起了人们对他所确立的动力学基本概念作精确的数学表述的巨大热情。

这一切,标志着欧洲自文艺复兴以来在资本主义生产力刺激下蓬勃发展的自然科学开始迈入突破阶段。同时,这种突破又面临着严重的数学上的困难。概括起来,主要表现在:

(1) 怎样确定运动物体的瞬时速度与加速度,以及反过来由速度或加速度公式确定物体移动的距离?

(2) 怎样作任意曲线的切线(如望远镜光程设计就涉及切线问题)?

(3) 怎样求函数的极大值、极小值(如行星与太阳间的最大和最小距离、抛射体的最大射程)?

(4) 怎样计算曲线长(如行星沿轨道运动的路程)、面积(如行星矢径扫过的面积)以及体积、重心与引力等。

要克服这些困难,无论是古希腊的几何还是中世纪及文艺复兴时代发展起来的代数,都明显地不能满足要求。因此,在17世纪上半叶,几乎所有的科学大师们都竭力寻求新的数学工具,特别是描述运动与变化的无限小算法,并且也是在牛顿诞生前后相当短的时期内,取得了迅速的进展,其中最重要的有:英国的耐普尔(John Naeipr,1550—1617)和瑞士的乔伯斯特(Jobst Bürgi,1552—1632)发明的对数(1614);费马(Pierre de Fermat,1601—1665)求极大值与极小值的方法(1629);卡瓦列里(Cavalieri,Francesco Bonaventura 1598—1647)的"不可分量原理"(1635);笛卡儿的解析几何(1637);英国17世纪仅次于牛顿的、最富独创性的数学家及瓦里斯(John Wallis,1616—1703)导入数学的无穷级数与圆周率的无穷乘积公式(1655),他1656年完成的《无穷算术》里,实质上提出了定积分的概念;等等。这一系列前驱者的工作,沿着不同的道路向近代数学的大门逼近,尤其是解析几何,把变量引进几何学,使运动与变化的定量表述成为可能。然而,解析几何本身还不是无限小算法,而上述费马等人对于求解具体的无限小问题的方法缺乏一般性,尚不能适应当时科学的普遍需要。因此,站在更高的高度,将以往分散的成果综合为统一的理论,是17世纪中叶自然科学赋予数学的紧迫任务。牛顿正是在这样的时刻出场。时代的需要和个人的胆识、才能,使他以最卓越的数学创造——微积分的发明而成为上述任务的完成者。

### 13.1.2　微积分在力学中的应用

**1. 导数和微分**

我们先简要地回顾一下微分学的两个基本概念——导数和微分。

在某一变化过程中,有两个相互联系着的变量 $x$ 和 $y$,$y$ 是 $x$ 的函数,并记为 $y=f(x)$。自变量 $x$ 由一个数值 $x_0$ 变到 $x_1$,其增量表示为 $\Delta x=x_1-x_0$,与之相应的因变量 $y$ 的数值将由 $y_0=f(x_0)$变到 $y_1=f(x_1)$,于是它的增量为 $\Delta y=y_1-y_0=f(x_1)-f(x_0)=f(x_0+\Delta x)-f(x_0)$,$\Delta y$ 与 $\Delta x$ 的比值即$\dfrac{\Delta y}{\Delta x}=\dfrac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$,叫作自变量由 $x=x_0$ 到 $x=$

$x_0+\Delta x$ 区间函数的平均变化率。$\Delta x$ 越小，$\frac{\Delta y}{\Delta x}$就越能精确地反映出函数 $y=f(x)$在 $x_0$ 这一点附近的变化率。然而，当 $\Delta x$ 趋近于零时，$\Delta y$ 也随着趋近于零，这时单纯考虑 $\Delta x$ 和 $\Delta y$ 都是没有意义的，因为都不是一个有限值，但它们的依赖的关系(比值)$\frac{\Delta y}{\Delta x}$在 $\Delta x$ 趋于零时往往具有确定的有限值(极限值)，在物理上，这个极限值在一定情况下就具有一定的意义。通常把 $\Delta x\to 0$ 时的比值$\frac{\Delta y}{\Delta x}$的极限值叫作函数 $y=f(x)$对 $x$ 的导数，并记作 $y'$或 $f'(x)$，即 $y'=f'(x)=\lim\limits_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim\limits_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$。这时我们也说函数 $y=f(x)$在 $x$ 处可导。导数反映出函数相对于自变量的变化快慢的程度，这个含义正是速度或加速度的物理意义。

在微分学中，我们把趋近于零的变量的增量 $\Delta x$ 表示为 $\mathrm{d}x$，而相应的函数 $y$ 的增量 $\Delta y$ 表示为 $\mathrm{d}y$，$\mathrm{d}x$ 叫作自变量的微分，$\mathrm{d}y$ 叫作函数 $y=f(x)$在 $x$ 处的微分。引入微分的概念以后，导数则可用微分的商来表示，即 $y'=f'(x)=\frac{\mathrm{d}y}{\mathrm{d}x}$。$f'(x)$叫作函数 $y=f(x)$对自变量 $x$ 的一阶导数。与此同时，把 $\mathrm{d}y=f(x)\mathrm{d}x$ 叫作函数对自变量 $x$ 的微分。

如果导数 $f'(x)$本身也是 $x$ 的函数，我们就可以再取它对 $x$ 的导数，这叫作函数 $y=f(x)$的二阶导数，记作 $y''$或 $f''(x)$、$\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$等，即 $y''=f''(x)=\frac{\mathrm{d}^2y}{\mathrm{d}x^2}=\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)=\frac{\mathrm{d}}{\mathrm{d}x}f'(x)$。

**2. 导数和微分在力学中的应用**

我们知道，在几何上，由导数的定义可知，函数 $y=f(x)$在 $x_0$ 处的导数 $f'(x_0)$表示曲线 $y=f(x)$在点$(x_0,f(x_0))$处的曲线的斜率。而在物理上它则可以用来表示各种具有不同意义的变量的变化“快慢”。直线和曲线运动中的速度和加速度，刚体转动中的角速度和角加速度以及谐振中的速度和加速度等物理量的关系式都是利用导数建立起来的。如在直线运动中，已知运动学方程 $x$，瞬时速度 $v$ 是运动学方程 $x$ 对时间 $t$ 的一阶导数，而瞬时加速度 $a$ 是瞬时速度 $v=v(t)$对时间 $t$ 的导数，亦即 $a$ 是 $x$ 对 $t$ 的二阶导数；瞬时功率的表达式 $N=\boldsymbol{F}\cdot\boldsymbol{v}$ 同样可以通过功的表达式 $A=\boldsymbol{F}\cdot\boldsymbol{r}$ 对时间的求导而得到。

通常，在我们的力学问题中，很多物理量如速度、力、功、动量、转动惯量等都不是常量，而是随着某一范围内的长度、角度、面积、体积、时间等而连续变化的变量。如对于处在液体中不同深度的物体表面(例如水坝的闸门等)来说，各处所承受的压强随着液体的深度的变化而有所不同，要计算物体表面所受的总压力时，一般就须用积分计算。再如作用在刚体上的力如果是连续分布的，则这些力对某些定点定轴的力矩代数和便可用积分求出。

在力学中，还常常遇到这么一类问题，在一定条件下，求物体运动的最大高度，最大射程和最短路程等。这一类的问题往往可以归结为求函数的最大值与最小值问题。我们知道，凡函数值由逐渐增大变为逐渐减小时，必经过一最大值。同样，凡函数值由逐渐减小变为逐渐增大时，必经过一最小值。无论对于最大值或最小值，曲线在该点的切线均成水平线，其斜率为零，由导数的几何意义可知，即 $\mathrm{d}y/\mathrm{d}x=0$，故求函数 $y=f(x)$的最大值或最小值时，首先只需求出它的一次导数 $\mathrm{d}y/\mathrm{d}x=f'(x)$，再解 $f'(x)=0$ 的方程，求其根 $x$ 的值。此外，我们还可以利用导数和微分得到牛顿第二定律公式以及推导出动量守恒定律。

积分是求导的逆运算。概括起来说，就是解决如下的数学问题：设函数 $y=f(x)$在区

间$[a,b]$上连续，把区间$[a,b]$分成$n$个小段，设每小段的大小为$\Delta x$，当$n\to 0$，即$\Delta x\to 0$时，$\sum_{i=1}^{n} f(x_i)\Delta x_i$的极限用符号$\int_a^b f(x)\mathrm{d}x$来表示，即$\lim\limits_{\Delta x\to 0}\sum_{i=1}^{n} f(x_i)\Delta x_i=\int_a^b f(x)\mathrm{d}x$。$\int_a^b f(x)\mathrm{d}x$叫作$x=a$到$x=b$区间内$f(x)$对$x$的定积分，而$f(x)$叫被积函数，$b$和$a$分别叫作定积分的上限和下限。定积分$\int_a^b f(x)\mathrm{d}x$在几何上表示曲线$y=f(x)$，两条直线$x=a$，$x=b$与$x$轴所围成的曲边梯形的面积。

**3. 导数和微分在力学中的应用方法和举例**

利用积分处理力学问题一般都按照如下的步骤。

(1) 建立坐标系，取自变量微元：把积分变量即自变量$x$的区间分为$n$个长度为$\mathrm{d}x$的小区间；如位移元大小$\mathrm{d}l$、面积元$\mathrm{d}s$、体积元$\mathrm{d}v$、时间元$\mathrm{d}t$等；

(2) 取函数微元：把待求量$y$的微分用$\mathrm{d}x$表示成被积表达式即：$\mathrm{d}y=f(x)\mathrm{d}x$，虽然被积函数$f(x)$是连续变化的，但在很短一段自变量区间$\mathrm{d}x$内，变化很小，近似不变。这一步是把具体的物理问题转换为数学问题，最关键。

(3) 求积分：在自变量区间内对$f(x)\mathrm{d}x$积分，即可求出$y$在该区间的和，$F(x)=\int f(x)\mathrm{d}x$。这一步不仅仅是单纯的数学计算，还要考虑初始(边界)条件，所以，必然牵涉物理问题。

(4) 讨论结果：对积分计算的结论赋予物理意义。

举例，设空气对抛体的阻力与抛体的速度成正比，即$\boldsymbol{F}_\mathrm{r}=-k\boldsymbol{v}$，$k$为比例系数，抛体的质量为$m$、初速为$\boldsymbol{v}_0$、抛射角为$\alpha$。求抛体运动的轨迹方程。

(1) 建立坐标系，取自变量微元：由$\boldsymbol{F}_\mathrm{r}=-k\boldsymbol{v}=-k(v_x\boldsymbol{i}+v_y\boldsymbol{j})$，把积分变量即自变量$v_x$的区间分为$n$个长度为$\mathrm{d}v_x$的小区间；自变量$v_y$的区间分为$n$个长度为$\mathrm{d}v_y$的小区间；$m\dfrac{\mathrm{d}v_x}{\mathrm{d}t}=-kv_x$；自变量$v_y$的区间分为$n$个长度为$\mathrm{d}v_y$的小区间；$m\dfrac{\mathrm{d}v_y}{\mathrm{d}t}=-mg-kv_y$；

(2) 取函数微元：把待求量$y$的微分用$\mathrm{d}x$表示成被积表达式，即：$\mathrm{d}y=f(x)\mathrm{d}x$。

$$\frac{\mathrm{d}v_x}{v_x}=-\frac{k}{m}\mathrm{d}t;\qquad \frac{k\,\mathrm{d}v_y}{mg+kv_y}=-\frac{k}{m}\mathrm{d}t$$

(3) 求积分：在自变量区间内对$f(x)\mathrm{d}x$积分，即可求出$y$在该区间的和。

方法一：可以使用不定积分，即$\begin{cases}\displaystyle\int\frac{\mathrm{d}v_x}{v_x}=-\int\frac{k}{m}\mathrm{d}t,\\ \displaystyle\int\frac{k\,\mathrm{d}v_y}{mg+kv_y}=-\int\frac{k}{m}\mathrm{d}t,\end{cases}$然后根据初始条件：$t=0$，速度在两个方向的分量分别为$v_{x0}=v_0\cos\alpha$和$v_{0y}=v_0\sin\alpha$，得最后结果：

$$v_x=v_0\cos\alpha\,\mathrm{e}^{-kt/m}\ \text{和}\ v_y=\left(v_0\sin\alpha+\frac{mg}{k}\right)\mathrm{e}^{-kt/m}-\frac{mg}{k}。$$

方法二：可以使用定积分，即$\begin{cases}\displaystyle\int_{v_{0x}}^{v_x}\frac{\mathrm{d}v_x}{v_x}=-\int_0^t\frac{k}{m}\mathrm{d}t,\\ \displaystyle\int_{v_{oy}}^{v_y}\frac{k\,\mathrm{d}v_y}{mg+kv_y}=-\int_0^t\frac{k}{m}\mathrm{d}t,\end{cases}$直接得$v_x=v_0\cos\alpha\,\mathrm{e}^{-kt/m}$

和 $v_y=\left(v_0\sin\alpha+\frac{mg}{k}\right)e^{-kt/m}-\frac{mg}{k}$。

(4) 讨论结果：对积分计算的结论赋予物理意义。

讨论 $v_y=\left(v_0\sin\alpha+\frac{mg}{k}\right)e^{-kt/m}-\frac{mg}{k}$ 和 $v_x=v_0\cos\alpha e^{-kt/m}$ 的物理意义。(略)

## 13.2 电磁学中的曲线和曲面积分

在普通物理学的电磁学里,无处不体现着高等数学的运用。场强、势能的定义和计算需要定积分；环流、通量的计算需要曲线、曲面积分等,其结论的含义往往超越了数学计算本身。本节从电磁学中的几个具有代表性的核心知识点出发,剖析高等数学中曲线、曲面积分在电磁学中的运用,进一步体会电磁学里的相关概念和原理的物理意义。

### 13.2.1 曲线积分

**1. 对弧长的曲线积分**

1) 对弧长的曲线积分的概念和性质

**定义** 设 $L$ 为 $xOy$ 面内的一条光滑的曲线弧,函数 $f(x,y)$在 $L$ 上有界。在 $L$ 上任意插入一点列 $M_1,M_2,\cdots,M_{n+1}$ 把 $L$ 分成 $n$ 个小段。设第 $i$ 个小段的长度为 $\Delta s_i$,又$(\xi_i,\eta_i)$为第 $i$ 个小段上任意取定的一个点,作乘积 $f(\xi_i,\eta_i)\Delta s_i$ $(i=1,2,\cdots,n)$,并作和 $\sum_{i=1}^{n}f(\xi_i,\eta_i)\Delta s_i$,如果当各小弧段的长度的最大值 $\lambda\to0$ 时,这和的极限总存在,则称此极限为函数 $f(x,y)$在曲线弧 $L$ 上对弧长的曲线积分或第一类曲线积分,记作 $\int_L f(x,y)\mathrm{d}s$,即

$$\int_L f(x,y)\mathrm{d}s=\lim_{i\to0}\sum_{i=1}^{n}f(\xi_i,\eta_i)\Delta s_i$$

其中,$f(x,y)$叫作被积函数,$L$ 叫作积分弧段。

如果 $L$ 是分段光滑的,我们规定函数在 $L$ 上的曲线积分等于函数在光滑的各段上的曲线积分之和。例如,设 $L$ 可分成两段光滑曲线弧 $L_1$ 和 $L_2$(记作 $L=L_1+L_2$),就规定

$$\int_L f(x,y)\mathrm{d}s=\int_{L_1}f(x,y)\mathrm{d}s+\int_{L_2}f(x,y)\mathrm{d}s$$

如果 $L$ 是闭曲线,那么函数 $f(x,y)$在闭曲线 $L$ 上对弧长的曲线积分记为 $\oint_L f(x,y)\mathrm{d}s$。

**性质** 由对弧长的曲线积分的定义可知,它有以下性质:

(1) $\int_L[f(x,y)\pm g(x,y)]\mathrm{d}s=\int_L f(x,y)\mathrm{d}s\pm\int_L g(x,y)\mathrm{d}s$。

(2) $\int_L kf(x,y)\mathrm{d}s=k\int_L f(x,y)\mathrm{d}s$ ($k$ 为常数)。

(3) $\int_L f(x,y)\mathrm{d}s=\int_{L_1}f(x,y)\mathrm{d}s+\int_{L_2}f(x,y)\mathrm{d}s$ $(L=L_1+L_2)$。

2) 对弧长的曲线积分的计算法

**定理** 设 $f(x,y)$在曲线弧 $L$ 上有定义且连续,$L$ 的参数方程为

$$\begin{cases} x = \phi(t) \\ y = \varphi(t) \end{cases} \quad \alpha \leqslant t \leqslant \beta$$

其中,$\varphi(t)$、$\phi(t)$在$[\alpha,\beta]$上具有一阶连续导数,且 $\varphi'^2(t)+\phi'^2(t)\neq 0$,则曲线积分 $\int_L f(x,y)\mathrm{d}s$ 存在,且

$$\int_L f(x,y)\mathrm{d}s = \int_\alpha^\beta f[\phi(t),\varphi(t)]\sqrt{\varphi'^2(t)+\phi'^2(t)}\,\mathrm{d}t \quad (\alpha < \beta)$$

**2. 对坐标的曲线积分**

1) 对坐标的曲线积分的概念和性质

**定义**:设 $L$ 为 $xOy$ 面内从点 $A$ 到点 $B$ 的一条有向光滑曲线弧,函数 $P(x,y)$、$Q(x,y)$在 $L$ 上有界,在 $L$ 上沿 $L$ 的方向任意插入一点列 $M_1(x_1,y_1),M_2(x_2,y_2),\cdots,M_{n-1}(x_{n-1},y_{n-1})$把 $L$ 分成 $n$ 个有向小弧段

$$M_{i-1}M_i(i=1,2,\cdots,n;\ M_0=A,M_n=B)$$

设 $\Delta x_i = x_i - x_{i-1}$,$\Delta y_i = y_i - y_{i-1}$,点$(\xi_i,\eta_i)$为 $M_{i-1}M_i$ 上任意取定的点。如果当各小弧段长度的最大值 $\lambda\to 0$ 时,$\sum\limits_{i-1}^{n} P(\xi_i,\eta_i)\Delta x_i$ 的极限总存在,则称此极限为函数 $P(x,y)$ 在有向曲线弧 $L$ 上对坐标 $x$ 的曲线积分,记作$\int_L P(x,y)\mathrm{d}x$。类似地,如果$\lim\limits_{\lambda\to 0}\sum\limits_{i=1}^{n} Q(\xi_i,\eta_i)\Delta y_i$ 总存在,则称此极限为函数 $Q(x,y)$ 在有向曲线弧 $L$ 上对坐标 $y$ 的曲线积分,记作 $\int_L Q(x,y)\mathrm{d}y$,即

$$\int_L P(x,y)\mathrm{d}x = \lim_{\lambda\to 0}\sum_{i=1}^{n} P(\xi_i,\eta_i)\Delta x_i$$

$$\int_L Q(x,y)\mathrm{d}y = \lim_{\lambda\to 0}\sum_{i=1}^{n} Q(\xi_i,\eta_i)\Delta y_i$$

其中 $P(x,y)$、$Q(x,y)$叫作被积函数,$L$ 叫作积分弧段。

以上两个积分也称为第二类曲线积分。应用上常出现的是

$$\int_L P(x,y)\mathrm{d}x + \int_L Q(x,y)\mathrm{d}y$$

这种合并起来的形式,为简便起见,把上式写成

$$\int_L (P(x,y)\mathrm{d}x + Q(x,y)\mathrm{d}y)$$

**性质**:根据上述曲线积分的定义,可以导出对坐标的曲线积分的如下性质:

(1) 如果把 $L$ 分成 $L_1$ 和 $L_2$,则

$$\int_L P\mathrm{d}x + Q\mathrm{d}y = \int_{L_1}(P\mathrm{d}x + Q\mathrm{d}y) + \int_{L_2}(P\mathrm{d}x + Q\mathrm{d}y)$$

(2) 设 $L$ 是有向曲线弧,$-L$ 是与 $L$ 方向相反的有向曲线弧,则

$$\int_{-L} P(x,y)\mathrm{d}x = -\int_L P(x,y)\mathrm{d}x$$

$$\int_{-L} Q(x,y)\mathrm{d}y = -\int_L Q(x,y)\mathrm{d}y$$

因此,关于坐标的曲线积分,我们必须注意积分弧段的方向。

2）对坐标的曲线积分的计算法

**定理** 设 $P(x,y)$、$Q(x,y)$在有向曲线弧 $L$ 上有定义且连续，$L$ 的参数方程为

$$\begin{cases} x=\varphi(t) \\ y=\phi(y) \end{cases}$$

当参数 $t$ 单调地由 $\alpha$ 变到 $\beta$ 时，点 $M(x,y)$从 $L$ 的起点 $A$ 沿 $L$ 运动到终点 $B$，$\varphi(t)$、$\phi(t)$在以 $\alpha$ 及 $\beta$ 为端点的闭区间上具有一阶连续导数，且 $\varphi'^2(t)+\phi'^2(t)\neq 0$，则曲线积分 $\int_L P(x,y)\mathrm{d}x+Q(x,y)\mathrm{d}y$ 存在，且

$$\int_L P(x,y)\mathrm{d}x+Q(x,y)\mathrm{d}y=\int_\alpha^\beta \{P[\varphi(t),\phi(t)]\varphi'(t)-Q[\varphi(t)-\phi(t)]\phi'(t)\}\mathrm{d}t$$

## 13.2.2 曲面积分

### 1. 对面积的曲面积分

1）对面积的曲面积分的概念与性质

**定义** 设曲面 $\Sigma$ 是光滑的，函数 $f(x,y,z)$在 $\Sigma$ 上有界。把 $\Sigma$ 任意分成 $n$ 小块 $\Delta S_i$（$\Delta S_i$ 同时也代表第 $i$ 小块曲面的面积），设$(\xi_i,\eta_i,\zeta_i)$是 $\Delta S_i$ 上任意取定的一点，作乘积 $f(\xi_i,\eta_i,\zeta_i)\Delta S_i(i=1,2,\cdots,n)$，并作和 $\sum\limits_{i=1}^{n} f(\xi_i,\eta_i,\zeta_i)\Delta S_i$。如果当时各小块曲面的直径的最大值 $\lambda\to 0$ 时，这和的极限总存在，则称此极限为函数 $f(x,y,z)$ 在曲面 $\Sigma$ 上对面积的曲面积分或第一类曲面积分，记作$\iint\limits_{\Sigma} f(x,y,z)\mathrm{d}s$，即

$$\iint_{\Sigma} f(x,y,z)\mathrm{d}s=\lim_{\lambda\to 0}\sum_{i=1}^{n} f(\xi_i,\mu_i,\zeta_i)\Delta S_i$$

其中 $f(x,y,z)$叫作被积函数，$\Sigma$ 叫作积分曲面。

如果 $\Sigma$ 是分片光滑的，我们规定函数在 $\Sigma$ 上对面积的曲面积分等于函数在光滑的各片曲面上对面积的曲面积分之和。例如，设 $\Sigma$ 可以分成两片光滑曲面 $\Sigma_1$ 和 $\Sigma_2$（记作 $\Sigma=\Sigma_1+\Sigma_2$），就规定

$$\iint_{\Sigma} f(x,y,z)\mathrm{d}s=\iint_{\Sigma_1} f(x,y,z)\mathrm{d}s+\iint_{\Sigma_2} f(x,y,z)\mathrm{d}s$$

由对面积的曲面积分的定义可知，它具有与对弧长的曲线积分相似的性质，这里就不再赘述了。

2）对面积的曲面积分的计算方法

设积分曲面 $\Sigma$ 由方程 $z=z(x,y)$给出，$\Sigma$ 在 $xOy$ 面上的投影区域为 $D_{xy}$，函数 $z=z(x,y)$在 $D_{xy}$ 上具有连续偏导数，被积函数 $f(x,y,z)$在 $\Sigma$ 上连续。则有

$$\begin{aligned}\iint_{\Sigma} f(x,y,z)\mathrm{d}s&=\lim_{\lambda\to 0}\sum_{i=1}^{n} f(\xi_i,\mu_i,\zeta_i)\Delta S_i \\ &=\iint_{D} f[x,y,z(x,y)]\sqrt{1+z_x^2+z_y^2}\,\mathrm{d}x\mathrm{d}y\end{aligned}$$

即可把对面积的曲面积分化为二重积分。

**2. 对坐标的曲面积分**

1）对坐标的曲面积分的概念与性质

**定义**　设Σ为光滑的有向曲面，函数$R(x,y,z)$在Σ上有界，把Σ任意分成$n$块小曲面$\Delta S_i$（$\Delta S_i$同时又表示第$i$块小曲面的面积），$\Delta S_i$在$xOy$面上的投影为$(\Delta S_i)_{xy}$，$(\xi_i,\eta_i,\zeta_i)$是$\Delta S_i$上任意取定的一点。如果当各小块曲面的直径的最大值$\lambda\to 0$时，

$$\lim_{\lambda\to 0}\sum_{i=1}^{n}R(\xi_i,\eta_i,\zeta_i)(\Delta S_i)_{xy}$$

总存在，则称此极限为函数$R(x,y,z)$在有向曲面Σ上对坐标$x$、$y$的曲面积分记作$\iint\limits_{\Sigma}R(x,y,z)\mathrm{d}x\mathrm{d}y$，即

$$\iint\limits_{\Sigma}R(x,y,z)\mathrm{d}x\mathrm{d}y=\lim_{\lambda\to 0}\sum_{i=1}^{n}R(\xi_i,\eta_i,\zeta_i)(\Delta S_i)_{xy}$$

其中$R(x,y,z)$叫作被积函数，Σ叫作积分曲面。

类似地，可以定义函数$P(x,y,z)$在有向曲面Σ上对坐标$y$、$z$的曲面积分$\iint\limits_{\Sigma}P(x,y,z)\mathrm{d}y\mathrm{d}z$，及函数$Q(x,y,z)$在有向曲面Σ上对坐标$z$、$x$的曲面积分$\iint\limits_{\Sigma}Q(x,y,z)\mathrm{d}z\mathrm{d}x$分别为

$$\iint\limits_{\Sigma}P(x,y,z)\mathrm{d}y\mathrm{d}z=\lim_{\lambda\to 0}\sum_{i=1}^{n}P(\xi_i,\eta_i,\zeta_i)(\Delta S_i)_{yz}$$

$$\iint\limits_{\Sigma}Q(x,y,z)\mathrm{d}z\mathrm{d}x=\lim_{\lambda\to 0}\sum_{i=1}^{n}Q(\xi_i,\eta_i,\zeta_i)(\Delta S_i)_{zx}$$

以上三个曲面积分也称为第二类曲面积分。

在应用上出现较多的是

$$\iint\limits_{\Sigma}R(x,y,z)\mathrm{d}x\mathrm{d}y+\iint\limits_{\Sigma}P(x,y,z)\mathrm{d}y\mathrm{d}z+\iint\limits_{\Sigma}Q(x,y,z)\mathrm{d}z\mathrm{d}x$$

这种合并起来的形式，为简便起见，我们把它写成

$$\iint\limits_{\Sigma}R(x,y,z)\mathrm{d}x\mathrm{d}y+P(x,y,z)\mathrm{d}y\mathrm{d}z+Q(x,y,z)\mathrm{d}z\mathrm{d}x$$

如果Σ是分片光滑的有向曲面，我们规定函数在Σ上对坐标的曲面积分等于函数在各片光滑曲面上对坐标的曲面积分之和。

2）对坐标的曲面积分的计算法

设积分曲面Σ由方程$z=z(x,y)$给出的曲面的上侧，Σ在$xOy$面上的投影区域为$D_{xy}$，函数$z=z(x,y)$在$D_{xy}$上具有一阶连续偏导数，被积函数$R(x,y,z)$在Σ上连续。则有

$$\iint\limits_{\Sigma}R(x,y,z)\mathrm{d}x\mathrm{d}y=\iint\limits_{D_{xy}}R[x,y,z(x,y)]\mathrm{d}x\mathrm{d}y$$

这就是把对坐标的曲面积分化为二重积分的公式。必须注意，上述公式的曲面积分取在曲面的上侧，如果曲面积分取在0的下侧，那么公式则变为

$$\iint_{\Sigma} R(x,y,z)\mathrm{d}x\mathrm{d}y = -\iint_{D_{xy}} R[x,y,z(x,y)]\mathrm{d}x\mathrm{d}y$$

类似地，$\iint_{\Sigma} P(x,y,z)\mathrm{d}y\mathrm{d}z$ 和$\iint_{\Sigma} Q(x,y,z)\mathrm{d}z\mathrm{d}x$ 计算方法也是一样的，这里不再赘述。

### 13.2.3 电磁学中曲线、曲面积分的运用

**1. 静止的点电荷的电场叠加**

我们先讨论一个静止的点电荷的场强分布(距静止电荷 $q$ 距离为 $r$ 的 $P$ 点处的场强)。设把一个检验电荷 $q_0$ 放在 $P$ 点，根据库仑定律，$q_0$ 受到的电场力为

$$\boldsymbol{F} = \frac{qq_0}{4\pi\varepsilon_0 r^2}\boldsymbol{e}_r \tag{13-2-1}$$

式中 $\boldsymbol{e}_r$ 是从场源电荷 $q$ 指向点 $P$ 的单位矢量。由场强定义式 $E=F/q_0$，$P$ 点的场强为

$$\boldsymbol{E} = \frac{q}{4\pi\varepsilon_0 r^2}\boldsymbol{e}_r \tag{13-2-2}$$

这就是点电荷场强分布公式。式中若 $q>0$，则 $\boldsymbol{E}$ 与 $\boldsymbol{r}$ 同向，即在正电荷周围的电场中，任意点的场强沿该点径矢方向；若 $q<0$，则 $\boldsymbol{E}$ 与 $\boldsymbol{r}$ 反向，即在负电荷周围的电场中，任意点的场强沿该点的径矢的反方向。

当电场由点电荷系 $q_1,q_2,\cdots,q_n$ 提供，其场强又如何求呢？由场强叠加原理可得，在 $n$ 个点电荷产生的电场中某点的电场强度等于每个点电荷单独存在时在该点所产生的电场强度的矢量和。因此，有

$$\boldsymbol{E} = \sum_{i=1}^{n} E_i = \sum_{i=1}^{n} \frac{q_i}{4\pi\varepsilon_0 r_i^2}\boldsymbol{e}_{r_i} \tag{13-2-3}$$

式中，$r_i$ 为 $q_i$ 到场点的距离，$\boldsymbol{e}_{r_i}$ 为从 $q_i$ 指向场点的单位矢量。

若带电体的电荷是连续分布的，可认为该带电体的电荷是由许多无穷小的电荷元 $\mathrm{d}q$ 组成的，而每个电荷元都可当作点电荷处理。设其中任一个电荷元 $\mathrm{d}q$ 在 $P$ 点产生的场强为 $\mathrm{d}\boldsymbol{E}$。按式(13-2-2)有

$$\mathrm{d}\boldsymbol{E} = \frac{\mathrm{d}q}{4\pi\varepsilon_0 r^2}\boldsymbol{e}_r \tag{13-2-4}$$

整个带电体在 $P$ 点所产生的总场强可用积分计算为

$$\boldsymbol{E} = \int \mathrm{d}\boldsymbol{E} = \int \frac{\mathrm{d}q}{4\pi\varepsilon_0 r^2}\boldsymbol{e}_r = \begin{cases} \displaystyle\int_L \frac{\lambda\,\mathrm{d}l}{4\pi\varepsilon_0 r^2}\boldsymbol{e}_r \\ \displaystyle\iint_S \frac{\sigma\,\mathrm{d}s}{4\pi\varepsilon_0 r^2}\boldsymbol{e}_r \\ \displaystyle\iiint_V \frac{\rho\,\mathrm{d}v}{4\pi\varepsilon_0 r^2}\boldsymbol{e} \end{cases} \tag{13-2-5}$$

式中 $r$ 是电荷元 $\mathrm{d}q$ 到场点 $P$ 的距离，而 $\boldsymbol{e}_r$ 是这一方向上的单位矢量，$\lambda$、$\sigma$ 和 $\rho$ 分别是线带电分布、面带电分布和体带电分布的电荷密度。

例如，如图 13-2-1 所示，一个细的带电塑料圆环，半径为 $R$，所带线电荷密度 $\lambda$ 与 $\theta$ 有 $\lambda=\lambda_0\sin\theta$ 的关系。试求圆心处的电场强度的大小和方向。

(1) 建立坐标系，取自变量微元：把积分变量即自变量 $x$ 的区间分为 $n$ 个长度为 $\mathrm{d}x$ 的

小区间。如图 13-2-1 所示，电荷元 $\mathrm{d}q=\lambda R\,\mathrm{d}\theta=\lambda_0\sin\theta R\,\mathrm{d}\theta$。

(2) 取函数微元：把待求量 $y$ 的微分用 $\mathrm{d}x$ 表示成被积表达式，即：$\mathrm{d}y=f(x)\mathrm{d}x$，虽然被积函数 $f(x)$ 是连续变化的，但在很短一段自变量区间 $\mathrm{d}x$ 内，变化很小，近似不变)。这一步是把具体的物理问题转换为数学问题，最关键。电荷元 $\mathrm{d}q$ 在圆心处的电场强度大小为

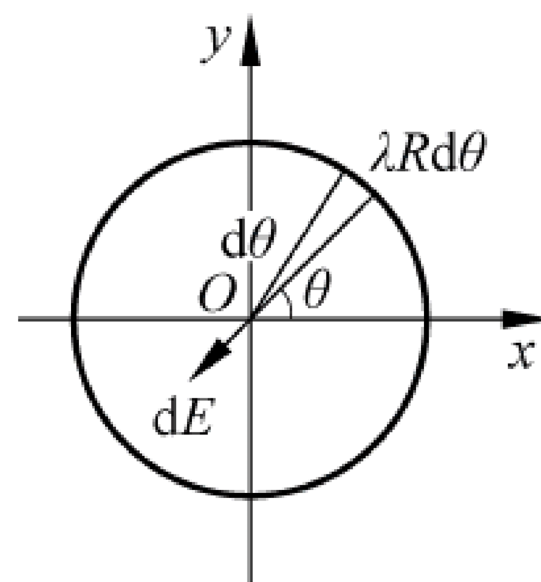



图 13-2-1 细的带电塑料圆环

$$\mathrm{d}E=\frac{\lambda_0\sin\theta\,\mathrm{d}\theta}{4\pi\varepsilon_0 R}$$

此电场强度的两个分量的函数微元为

$$\mathrm{d}E_x=\mathrm{d}E\cos\theta=\frac{\lambda_0\sin\theta\cos\theta\,\mathrm{d}\theta}{4\pi\varepsilon_0 R}$$

$$\mathrm{d}E_y=-\mathrm{d}E\sin\theta=-\frac{\lambda_0\sin^2\theta\,\mathrm{d}\theta}{4\pi\varepsilon_0 R}$$

(3) 求积分：在自变量区间内对 $f(x)\mathrm{d}x$ 积分，即可求出 $y$ 在该区间的和。将此二分量对 $\theta$ 从 $0-2\pi$ 积分，可得

$$E_x=\int \mathrm{d}E_x=0$$

$$E_y=\int \mathrm{d}E_y=\int_0^{2\pi}-\frac{\lambda_0\sin^2\theta\,\mathrm{d}\theta}{4\pi\varepsilon_0 R}=\frac{-\lambda_0}{4\varepsilon_0 R}$$

(4) 讨论结果：对积分计算的结论赋予物理意义。

最后得：$\boldsymbol{E}=E_x\boldsymbol{i}+E_y\boldsymbol{j}=-\dfrac{\lambda_0}{4\varepsilon_0 R}\boldsymbol{j}$

**2. 电通量**

如图 13-2-2 所示，以 $\mathrm{d}S$ 表示电场中某一个设想的面元。通过此面元的电场线条数就定义为通过这一面元的电通量。为了求出这一电通量，我们考虑此面元在垂直于场强方向的投影 $\mathrm{d}S_\perp$。很明显通过 $\mathrm{d}S$ 和 $\mathrm{d}S_\perp$ 的电场线条数是一样的。由图可知，$\mathrm{d}S_\perp=\mathrm{d}S\cos\theta$。可得通过 $\mathrm{d}S$ 的电场线条数或电通量应为

$$\mathrm{d}\Phi_e=E\,\mathrm{d}S_\perp=E\,\mathrm{d}S\cos\theta \tag{13-2-6}$$

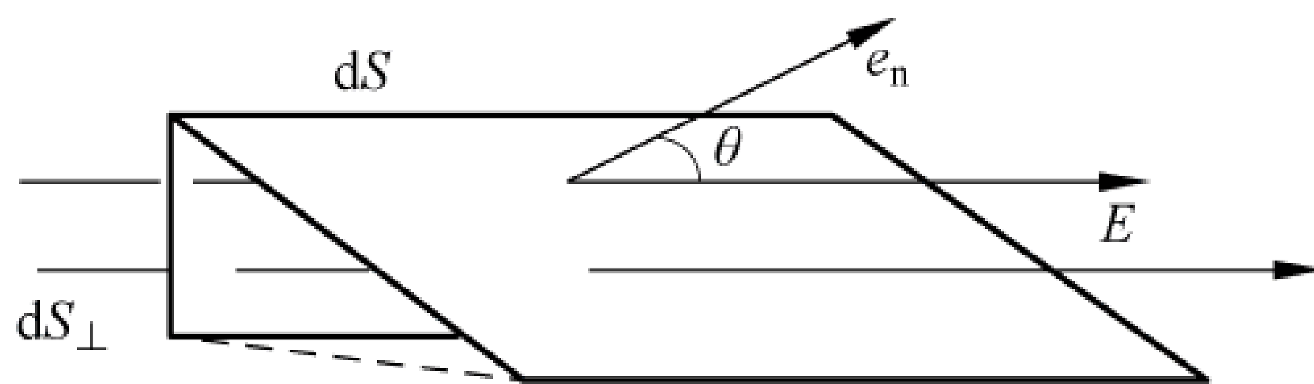



图 13-2-2 通过 dS 的电通量

为了同时表示出面元的方位，我们利用面元的法向单位矢量 $\boldsymbol{e}_n$，由矢量标积的定义，可得

$$\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{S}=\boldsymbol{E}\cdot\boldsymbol{e}_n\mathrm{d}S=E\,\mathrm{d}S\cos\theta \tag{13-2-7}$$

将式(13-2-6)和式(13-2-7)对比，可用矢量标积表示通过面元 dS 的电通量的公式

$$\mathrm{d}\Phi_e=\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{S} \tag{13-2-8}$$

注意，由式(13-2-8)决定的电通量 $\mathrm{d}\Phi_e$ 有正负之别。当 $0\leqslant\theta\leqslant\pi/2$ 时，$\mathrm{d}\Phi_e$ 为正；当 $\pi/2\leqslant\theta\leqslant\pi$ 时，$\mathrm{d}\Phi_e$ 为负。

为了求出通过任意曲面 $S$ 的电通量,可将曲面 $S$ 分割成许多小面元 $\mathrm{d}S$。先计算通过每一小面元的电通量,然后对整个 $S$ 面上所有面元的电通量相加。用数学式表示就有

$$\Phi_e=\int \mathrm{d}\Phi_e=\int_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} \tag{13-2-9}$$

此积分即为面积分,积分号下标 $S$ 表示积分遍及整个曲面。

通过一个封闭曲面 $S$ 的电通量可表示为

$$\Phi_e=\oint_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} \tag{13-2-10}$$

积分符号“$\oint$”表示对整个封闭曲面进行面积分。

**3. 高斯定律**

高斯(K. F. Gauss,1777—1855)是德国物理学家和数学家,他在实验物理和理论物理及数学方面都做出了很多贡献,他导出的高斯定律是电磁学中的一条重要定律。

高斯定律是用电通量表示的电场和场源电荷关系的定律,它给出了通过任意一个封闭面的电通量与封闭面内所包围的电荷的关系。下面我们利用电通量的概念根据库仑定律和场强叠加原理来导出这个关系。

先讨论静止的点电荷 $q$ 的电场。以 $q$ 所在点为中心,取任意长度 $r$ 为半径作一球面 $S$ 包围这个点电荷 $q$,如图 13-2-3 所示。

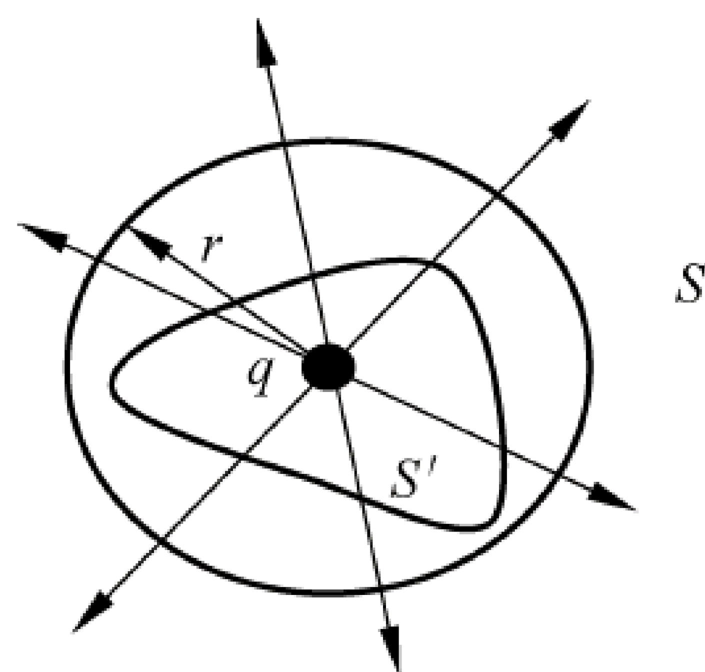


图 13-2-3 通过封闭面的电通量

球面上任意一点的电场强度 $\boldsymbol{E}$ 的大小都是$\dfrac{q}{4\pi\varepsilon_0 r^2}$,方向都沿着径矢 $\boldsymbol{r}$ 的方向,而且处处与球面垂直。可得通过这球面的电通量为

$$\Phi_e=\oint_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S}=\oint_S \frac{q}{4\pi\varepsilon_0 r^2}\mathrm{d}S=\frac{q}{4\pi\varepsilon_0 r^2}\oint_S \mathrm{d}S=\frac{q}{\varepsilon_0}$$

此结果与球面半径无关,只和它所包围的电荷的电量有关。这意味着,对以点电荷 $q$ 为中心的任意球面来说,通过它的电通量都一样,都为$\dfrac{q}{\varepsilon_0}$。

现设想另一个任意的闭合面 $S'$,$S'$与球面 $S$ 包围同一点电荷 $q$,如图 13-2-3 所示,由于电场线的连续性,可以得出通过闭合面 $S'$与 $S$ 的电力线数目是一样的。因此,通过任意形状的包围点电荷 $q$ 的闭合面的电通量都等于$\dfrac{q}{\varepsilon_0}$。

以上是关于单个点电荷的电场的讨论。对于一个由点电荷 $q_1,q_2,\cdots,q_n$ 等组成的点电荷系来说,有电场叠加可得

$$\Phi_e=\oint_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S}=\frac{1}{\varepsilon_0}\sum q_{\text{int}}$$

式中 $\sum q_{\text{int}}$ 表示在封闭曲面内的电量的代数和。此即高斯定律的数学表达式。它表明:在真空中的静电场内,通过任意封闭曲面的电通量等于该封闭曲面所包围的电荷电量的代数和的$\dfrac{1}{\varepsilon_0}$倍。

**4. 静电场的保守性**

运用曲线积分,从库仑定律出发可以证明静电场的保守性。

图 13-2-4 中，以 $q$ 表示固定于某处的一个点电荷，当另一电荷 $q_0$ 在它的电场中由 $P_1$ 点沿任一路径移到 $P_2$ 点时，$q_0$ 受的电场力做的功为

$$A_{12}=\int_{(P_1)}^{(P_2)}\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}=q_0\int_{(P_1)}^{(P_2)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}\qquad(13\text{-}2\text{-}11)$$

上式两侧除以 $q_0$，得到

$$\frac{A_{12}}{q_0}=\int_{(P_1)}^{(P_2)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}$$

图 13-2-4　电场力沿任一路径做的功

上式等号右侧的积分 $\int_{(P_1)}^{(P_2)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}$ 叫作电场强度 $\boldsymbol{E}$ 沿任意路径 $L$ 的线积分，它表示在电场中从 $P_1$ 点到 $P_2$ 点移动单位正电荷时电场力所做的功。由于这一积分只由 $q$ 的电场强度 $\boldsymbol{E}$ 的分布决定，而与被移动的电荷的电量无关，所以可以用它来说明电场的性质。

对于静止的点电荷 $q$ 来说，其电场强度公式为 $\boldsymbol{E}=\frac{q}{4\pi\varepsilon_0 r^2}\boldsymbol{e}_r=\frac{q}{4\pi\varepsilon_0 r^3}\boldsymbol{r}$，将此式代入式(13-2-11)中，得场强 $E$ 的线积分为

$$\int_{(P_1)}^{(P_2)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}=\int_{(P_1)}^{(P_2)}\frac{q}{4\pi\varepsilon_0 r^3}\boldsymbol{r}\cdot\mathrm{d}\boldsymbol{r}\qquad(13\text{-}2\text{-}12)$$

从图 13-2-4 可以看出，$\boldsymbol{r}\cdot\mathrm{d}\boldsymbol{r}=r\cos\theta|\mathrm{d}\boldsymbol{r}|=r\mathrm{d}r$，这里 $\theta$ 是从电荷 $q$ 引到 $q_0$ 的径矢与 $q_0$ 的位移元 $\mathrm{d}\boldsymbol{r}$ 之间的夹角。将此关系代入上式，得

$$\int_{(P_1)}^{(P_2)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}=\int_{r_1}^{r_2}\frac{q}{4\pi\varepsilon_0 r^2}\mathrm{d}r=\frac{q}{4\pi\varepsilon_0}\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

由于 $r_1$ 和 $r_2$ 分别表示从点电荷 $q$ 到起点和终点的距离，所以此结果说明，在静止的点电荷 $q$ 的电场中，电场强度的积分只和积分路径的起点和终点的位置有关，而与积分路径无关。也可以说在静止的点电荷的电场中，移动单位正电荷时，电场力所做的功只取决于被移动电荷的起点和终点的位置，而与移动路径无关。

对于由许多静止的点电荷 $q_1,q_2,\cdots,q_n$ 组成的点电荷系，由场强叠加原理可知其总电场强度的线积分也具有这一特点。

对于静止的连续的带电体，可将其看作无数点电荷的集合，因而它的电场的强度的线积分同样具有这样的特点。

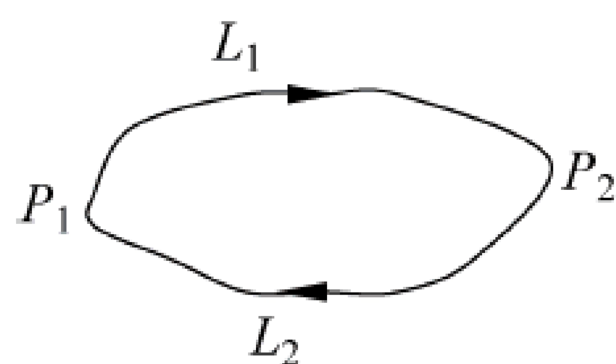


图 13-2-5　电场力沿闭合路径做的功

因此，我们可以得出结论，对任何静电场，电场强度的线积分 $\int_{(P_1)}^{(P_2)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}$ 都只取决于起点 $P_1$ 和终点 $P_2$ 的位置而与连接 $P_1$ 和 $P_2$ 点间的路径无关，静电场的这一特性叫作静电场的保守性。

静电场的保守性还可以表述成另一种形式。如图 13-2-5 所示，在静电场中作一任意闭合路径 $L$，考虑场强 $E$ 沿此闭合路径的线积分。在 $L$ 上取任意两点 $P_1$ 和 $P_2$，它们把 $L$ 分成 $L_1$ 和 $L_2$ 两段，因此，沿 $L$ 环路的场强的线积分为

$$\oint_L\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}={}_{L_1}\int_{(P_1)}^{(P_2)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}+{}_{L_2}\int_{(P_2)}^{(P_1)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}={}_{L_1}\int_{(P_1)}^{(P_2)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}-{}_{L_2}\int_{(P_1)}^{(P_2)}\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{r}$$

由于场强的线积分与路径无关，所以上式最后的两个积分值相等。因此

$$\oint_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{r} = 0 \tag{13-2-13}$$

此式表明,在静电场中,场强沿任意闭合路径的线积分等于零。这是静电场保守性的另一种说法,称作静电场环路定理。

**5. 电势叠加原理**

我们知道在真空中静止的电荷分布求其电场的电势分布时,除了直接用公式 $\varphi = \int_P^{(P_0)} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{r}$ 以外,还可以在点电荷电势公式 $\varphi = \dfrac{q}{4\pi\varepsilon_0 r}$ 的基础上应用叠加原理来求出结果。这后一方法的原理如下。

设场源电荷系由若干个带电体组成,由场强叠加原理,很容易得出:$\varphi = \sum \varphi_i$。此式称作电场叠加原理。它表示一个电荷系的电场中任一点的电势等于每一个带电体单独存在时在该点产生的电势的代数和。

实际上应用电势叠加原理时,可以从点电荷的电势出发,先考虑场源电荷系由许多点电荷组成的情况。这时将点电荷的电势公式

$$\varphi = \frac{q}{4\pi\varepsilon_0 r} \tag{13-2-14}$$

代入上式,可得点电荷系的电场中某点的电势为:

$$\varphi = \sum \frac{q_i}{4\pi\varepsilon_0 r_i} \tag{13-2-14$'$}$$

式中 $r_i$ 为从点电荷 $q_i$ 到该点的距离。对一个电荷连续分布的带电体,可以设想它由许多点电荷元 $\mathrm{d}q$ 组成。将每个电荷元都当成点电荷,就可以由上式得出用叠加原理求电势的积分公式

$$\varphi = \int \frac{\mathrm{d}q}{4\pi\varepsilon_0 r} \tag{13-2-15}$$

**6. 束缚电荷与电极化强度的关系**

由于电介质的束缚电荷是电介质极化的结果,所以束缚电荷和电极化强度之间一定存在某种定量关系,这一关系可如下求得。

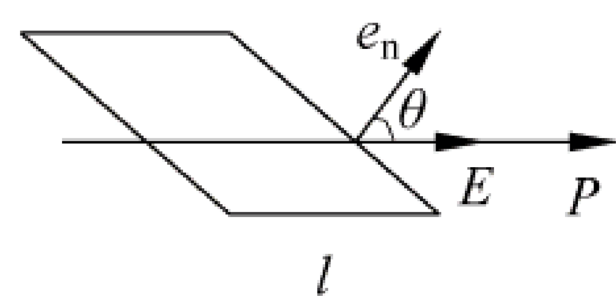


图 13-2-6 束缚电荷和电极化强度

以非极性电介质为例,考虑电介质内部某一小面元 $\mathrm{d}S$ 处的电极化。如图 13-2-6 所示,设电场 $\boldsymbol{E}$ 的方向(因而电极化强度矢量 $\boldsymbol{P}$ 的方向)和 $\mathrm{d}\boldsymbol{S}$ 的正法线方向 $\boldsymbol{e}_n$ 成 $\theta$ 角。为简单起见,假定负电荷不动,而正电荷沿 $\boldsymbol{E}$ 的方向发生位移 $l$。在面元 $\mathrm{d}\boldsymbol{S}$ 后侧取一斜高为 $l$,底面积为 $\mathrm{d}S$ 的体积元 $\mathrm{d}V$。

由于电场 $\boldsymbol{E}$ 的作用,此体积内所有分子的正电荷重心将越过 $\mathrm{d}\boldsymbol{S}$ 到前侧去。以 $q$ 表示每个分子的正电荷量,以 $n$ 表示电介质单位体积内的分子数,则由于电极化而越过 $\mathrm{d}S$ 面的总电荷为

$$\mathrm{d}q' = qn\mathrm{d}V = qnl\mathrm{d}S\cos\theta$$

由于 $ql = p$,而 $np = P$,所以

$$\mathrm{d}q' = P\cos\theta\mathrm{d}S$$

因此,$\mathrm{d}S$ 面上因电极化而越过单位面积的电荷为

$$\frac{\mathrm{d}q'}{\mathrm{d}S} = P\cos\theta = \boldsymbol{P} \cdot \boldsymbol{e}_n$$

这一关系式虽然是利用非极性分子电介质推出，但对极性分子电介质同样适用。

在上述论证中，如果 d$\boldsymbol{S}$ 面恰巧是电介质的面临真空的表面，而 $\boldsymbol{e}_n$ 是其外法线方向的单位矢量，则上式就给出因电极化而在电介质表面单位面积上显露出的面束缚电荷，即面束缚电荷密度。以 $\sigma'$ 表示面束缚电荷密度，则上述可得

$$\sigma' = P\cos\theta = \boldsymbol{P}\cdot\boldsymbol{e}_n \tag{13-2-16}$$

电介质内部体束缚电荷的产生可以根据式(13-2-16)进一步求出。为此可设想电介质内部任一封闭曲面 $S$，如图 13-2-7 所示，如上已求得由于电极化而越过 d$\boldsymbol{S}$ 面向外移出封闭面的电荷为 $\mathrm{d}q'_{\text{out}} = P\cos\theta\mathrm{d}S = \boldsymbol{P}\cdot\mathrm{d}\boldsymbol{S}$。通过整个封闭面向外移出的电荷应为

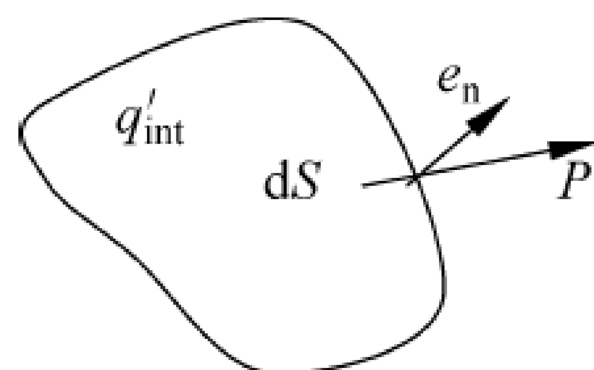


图 13-2-7　束缚电荷密度和电极化强度

$$q'_{\text{out}} = \oint_s \mathrm{d}q'_{\text{out}} = \oint_s \boldsymbol{P}\cdot\mathrm{d}\boldsymbol{S}$$

因为电介质是中性的，根据电荷守恒，由于电极化而在封闭面内留下的多余电荷，即体束缚电荷应为

$$q'_{\text{int}} = -q'_{\text{out}} = -\oint_S \boldsymbol{P}\cdot\mathrm{d}\boldsymbol{S} \tag{13-2-17}$$

这就是由于电介质极化而产生的体束缚电荷与电极化强度的关系：封闭面内的体束缚电荷等于通过该封闭面的电极化强度通量的负值。

**7. *D* 的高斯定律**

如图 13-2-8 所示，带电的导体和电极化了的电解质组成的系统可视为由一定的束缚电荷 $q'_{\text{int}}$ 和自由电荷 $q_o(\sigma_0)$ 分布组成的电荷系统，所有这些电荷产生一电场分布 $\boldsymbol{E}$。由高斯定律可知，对封闭面 $S$ 来说，

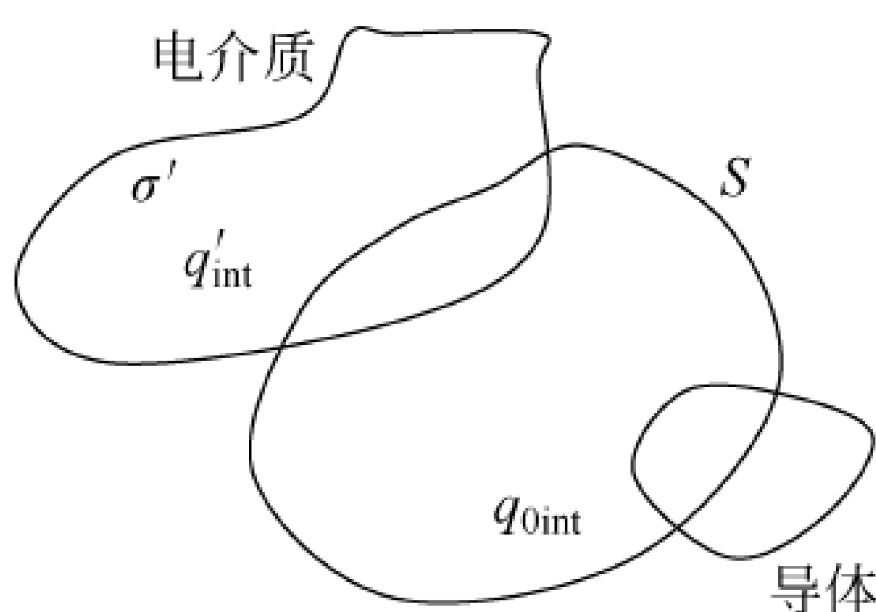


图 13-2-8　任意空间的高斯定律

$$\oint_S \boldsymbol{E}\cdot\mathrm{d}\boldsymbol{S} = \frac{1}{\varepsilon_0}\left(\sum q_{0\text{int}} + q'_{\text{int}}\right) \tag{13-2-18}$$

将 $q'_{\text{int}} = -q'_{\text{out}} = -\oint_S \boldsymbol{P}\cdot\mathrm{d}\boldsymbol{S}$ 式中的 $q'_{\text{int}}$ 代入此式，移项后可得，$\oint_S(\varepsilon_0\boldsymbol{E}+\boldsymbol{P})\cdot\mathrm{d}\boldsymbol{S} = \sum q_{0\text{int}}$。在此，引入一个辅助物理量——电位移——表示积分号里的合矢量，并以 $\boldsymbol{D}$ 表示，即定义

$$\boldsymbol{D} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P} \tag{13-2-19}$$

则上式就可简洁地表示为

$$\oint_S \boldsymbol{D}\cdot\mathrm{d}S = \sum q_{0\text{int}} \tag{13-2-20}$$

式(13-2-8)说明，通过任意封闭曲面的电位移通量等于该封闭面包围的自由电荷的代数和。这一关系叫作 $\boldsymbol{D}$ 的高斯定律，是电磁学的一条基本定律。

**8. 载流导线在磁场中受的力**

导线中的电流是由其中的载流子定向移动形成的，当把载流导线置于磁场中时，这些运动的载流子就要受到洛伦兹力的作用。其结果将表现为载流导线受到磁力的作用。为了计算一段载流导线受的磁力，先考虑它的一段长度元受的作用力。

设导线截面积为 $S$，其中有电流 $I$ 通过。考虑长度为 d$l$ 的一段导线，把它视为矢量，它的方向与电流的方向，即 $q\boldsymbol{v}$ 的方向相同。载有电流的一段矢量线元通常叫电流元。设导

线的单位体积内有 $n$ 个载流子，每一个载流子的电荷都是 $q$。我们认为各载流子都以漂移速度 $\boldsymbol{v}$ 运动。所以这些载流子受的力的总和就是

$$\mathrm{d}\boldsymbol{F}=nS\mathrm{d}lq\boldsymbol{v}\times\boldsymbol{B} \tag{13-2-21}$$

由于 $q\boldsymbol{v}$ 的方向和 $\mathrm{d}\boldsymbol{l}$ 的方向相同，所以 $q\mathrm{d}l\boldsymbol{v}=|q|v\mathrm{d}\boldsymbol{l}$。利用这一关系，式(13-2-21)就可写成 $\mathrm{d}\boldsymbol{F}=nSv|q|\mathrm{d}\boldsymbol{l}\times\boldsymbol{B}$，又由于 $nSv|q|=I$，即通过 $\mathrm{d}l$ 的电流强度的大小，所以最后可得

$$\mathrm{d}\boldsymbol{F}=I\mathrm{d}\boldsymbol{l}\times\boldsymbol{B} \tag{13-2-22}$$

$\mathrm{d}l$ 中的载流子由于受到这些力，所增加的动量最终总要传给导线本体的正离子结构，所以这一公式也就给出了这一段电流元受到的磁力。载流导线受磁场的作用力通常叫作安培力。

知道了一段电流元受的磁力就可以用积分的方法求出一段有限长导线 $L$ 受的磁力，如

$$\boldsymbol{F}=\int_L I\mathrm{d}\boldsymbol{l}\times\boldsymbol{B} \tag{13-2-23}$$

式(13-2-23)中 $\boldsymbol{B}$ 为各电流元所在处的“当地 $\boldsymbol{B}$”。

**9. 毕奥-萨伐尔定律**

电流在其周围产生磁场，其规律的基本形式是电流元产生的磁场和该电流元的关系。如图 13-2-9 所示，以 $I\mathrm{d}\boldsymbol{l}$ 表示稳恒电流的一电流元，以 $\boldsymbol{r}$ 表示从此电流元指向某一场点 $P$ 的径矢，实验指出，此电流元在 $P$ 点产生的磁场 $\mathrm{d}\boldsymbol{B}$ 由下式决定：

$$\mathrm{d}\boldsymbol{B}=\frac{\mu_0}{4\pi}\frac{I\mathrm{d}\boldsymbol{l}\times\boldsymbol{e}_r}{r^2} \tag{13-2-24}$$

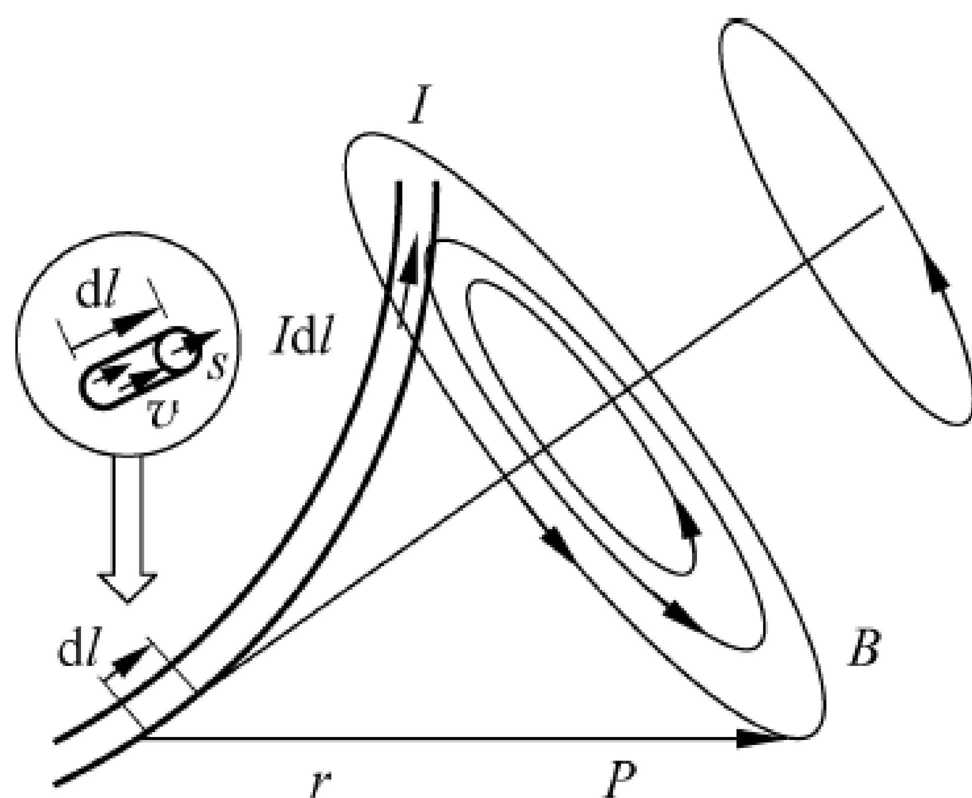



图 13-2-9 电流元的磁场

有了电流元的磁场公式，根据叠加原理，对这一公式进行积分，就可以求出任意电流的磁场分布。

根据式中的矢量关系可知，电流元的磁场的磁感线也都是圆心在电流元轴线上的同心圆。由于这些圆都是闭合曲线，所以通过任意封闭曲面的磁通量等于零。又由于任何电流都是一段段电流元组成的，根据叠加原理，在它的磁场中通过一个封闭曲面的磁通量应是各个电流元的磁场通过该封闭曲面的磁通量的代数和。既然每一个电流元的磁场通过该封闭曲面的磁通量为零，所以在任何磁场中通过任意封闭曲面的磁通量总等于零。这个关于磁场的结论叫作磁通连续定理，或磁场的高斯定理。其数学表达式为

$$\oint_S\boldsymbol{B}\cdot\mathrm{d}\boldsymbol{S}=0 \tag{13-2-25}$$

和电场的高斯定理相比，可知磁通连续定理反映自然界中没有与电荷相对应的“磁荷”(或叫单独的磁极或磁单极子)存在。近代关于基本粒子的理论研究早已预言有磁单极子存在，也曾企图在实验中找到它。但至今除了个别事件可作为例证外，还不能说完全肯定地发现了它。

**10. $\boldsymbol{H}$ 的环路定律**

磁介质在磁场中时，磁介质受磁场的作用要产生束缚电流，这束缚电流反过来又会影响磁场的分布。这时任一点的磁感应强度 $\boldsymbol{B}$ 应是自由电流的磁场 $\boldsymbol{B}_0$ 和束缚电流的磁场 $\boldsymbol{B}'$ 的矢量和，即 $\boldsymbol{B}=\boldsymbol{B}_0+\boldsymbol{B}'$。

由于束缚电流和磁介质磁化的程度有关，而这磁化的程度取决于磁感应强度 $\boldsymbol{B}$，所以磁介质和磁场的相互影响呈现一种比较复杂的关系。这样复杂关系也可以像研究电介质和电场的相互影响那样，通过引入适当的物理量而加以简化。下面就通过安培环路定理来导出

这种简化式。

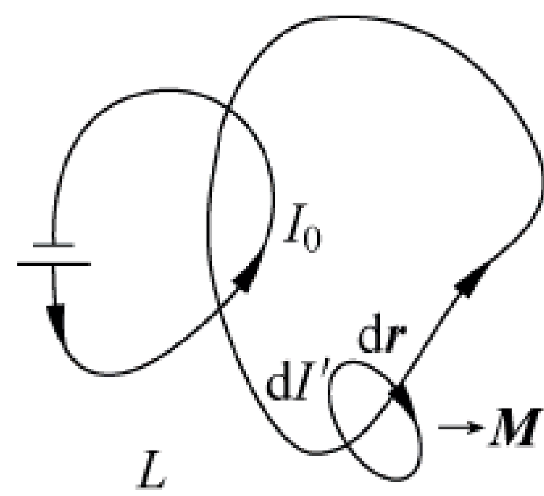


图 13-2-10　$H$ 的环路定理

载流导体和磁化了的磁介质组成的系统可视为由一定的自由电流 $I_0$ 和束缚电流 $I'(j')$ 分布组成的电流系统，如图 13-2-10 所示。所有这些电流产生一磁场分布 $\boldsymbol{B}$，由安培环路定理

$$\int_L \boldsymbol{B}\cdot \mathrm{d}\boldsymbol{r}=\mu_0\sum I_{\text{int}}$$

可知，对任一闭合路径 $L$，有 $\oint_L \boldsymbol{B}\cdot \mathrm{d}\boldsymbol{r}=\mu_0\sum(I_{0\text{int}}+I'_{\text{int}})$。将式 $I'=\oint_L \mathrm{d}I'=\oint_L \boldsymbol{M}\cdot \mathrm{d}\boldsymbol{r}$ 的 $I'$ 代入此式中的 $I'_{\text{int}}$，移项后可得

$$\oint_L\left(\frac{\boldsymbol{B}}{\mu_0}-\boldsymbol{M}\right)\cdot \mathrm{d}\boldsymbol{r}=\sum I_{0\text{int}} \tag{13-2-26}$$

在此，引入一辅助物理量表示积分号内的合矢量，叫作磁场强度，并以 $\boldsymbol{H}$ 表示，即定义

$$\boldsymbol{H}=\frac{\boldsymbol{B}}{\mu_0}-\boldsymbol{M} \tag{13-2-27}$$

则式(13-2-26)可简洁地表示为

$$\oint_L \boldsymbol{H}\cdot \mathrm{d}\boldsymbol{r}=\sum I_{0\text{int}} \tag{13-2-28}$$

式(13-2-28)说明，沿任一闭合路径磁场强度的环路积分等于该闭合路径所包围的自由电流的代数和。这一关系叫作 $\boldsymbol{H}$ 的环路定理，也是电磁学的一条基本定律。

## 13.3　热学中的统计方法

玻耳兹曼(Boltzmann，Ludwig，1844—1906)和麦克斯韦分别提出了研究宏观平衡性质的概率统计法，为统计力学奠定了基石。1902 年吉布斯出版了他的统计物理学书，把玻耳兹曼和麦克斯韦新创立的统计方法推广而发展成为系统的理论。气体分子运动论、统计力学和涨落现象理论是统计物理学三个组成部分。

### 13.3.1　统计方法简介

**1. 概率**

在一定条件下进行的实验中，一事件可能发生也可能不发生，必然发生的事件叫作必然事件，不可能发生的事件叫不可能事件。在重复实验次数极大时，随机事件出现的频率逼近的数称为在此条件下这一事件的概率。显然，必然事件的概率为 1，不可能事件的概率为 0，而随机事件的概率介于 0 与 1 之间。用数学式子表达就是：设在一定条件下，重复实验 $N$ 次，其中事件 $A$ 出现 $\Delta N$ 次，若 $N\to\infty$时，$A$ 出现的概率

$$P_A=\lim_{N\to\infty}\frac{\Delta N}{N} \tag{13-3-1}$$

概率具有加法和乘法定理。加法定理：几个互斥事件出现其中任一事件的概率为每个事件单独出现概率之和。互斥事件是指不能同时出现的随机事件。数学表达式为 $P_A=\sum_{i=1}^{n}P_{A_i}$。乘法定理：几个独立事件同时发生的概率等于各独立事件的概率之积。独立事件是指互不影响的事件。数学表达式为：$P_A=P_{A_1}P_{A_2}\cdots P_{A_n}$。

**2. 统计规律性**

统计规律性是指在大量随机事件的集合中出现的规律性，其具有以下两个特点：第一，

它是大量随机事件的集体表现,事件数量越多,规律性越稳定明显。第二,随着事件数量增加,频率趋向于确定的概率;概率是统计规律的主要特征。

**3. 统计平均值**

若随机物理量 $Q$ 是离散的,在 $\Delta N_1$ 个事件中取值 $Q_1$,在 $\Delta N_2$ 个事件中取值 $Q_2$,……,在 $\Delta N_i$ 个事件中取值 $Q_i$,……,在 $\Delta N_m$ 个事件中取值 $Q_m$;总事件数 $N=\sum_{i=1}^{m}\Delta N_i$。有 $\lim_{N\to\infty}\frac{\Delta N_i}{N}=P_i$,$P_i$ 就是 $Q$ 取值为 $Q_i$ 的概率。所以 $Q$ 的统计平均值

$$\bar{Q}=\lim_{N\to\infty}\sum_{i=1}^{m}Q_i\frac{\Delta N_i}{N}=\sum_{i=1}^{m}Q_iP_i \tag{13-3-2}$$

假若物理量 $Q$ 是连续变化的,则

$$\bar{Q}=\int_{N\to\infty}Q\frac{\mathrm{d}N}{N}=\int Q\mathrm{d}p \tag{13-3-3}$$

**4. 涨落与统计规律**

一可观测物理量某次测量结果与其相应的统计平均值的偏差,称为涨落。利用统计理论可以证明:如果事件发生的次数为 $N$,则实测值的涨落的最大幅度为$\pm\sqrt{N}$。

## 13.3.2 热力学系统统计描述的基本概念

**1. 等概率原理**

在热力学系统中,我们不能事先断定分子代表点要以多大的概率占据不同的相格,因此一种微观状态出现的概率是不能由直接计算得到的,即使是对于属于同一宏观态的各微观态也不能事先做出任何倾向性的论断。鉴于此,玻耳兹曼在19世纪70年代提出了著名的等概率原理:处于平衡态的孤立系统,其各个可能的微观状态出现的概率是相等的。这是统计物理中的一个基本假设,等概率原理的种种推论都与客观实际相符,而使其正确性得到了证实,其成为平衡态经典统计物理的基础。

**2. 热力学概率**

在统计物理中,给定分布下所对应的微观状态数目是可以计算的,以 $W_i$ 表示第 $i$ 种分布所对应的微观状态数。对于给定的孤立系统,考虑到各种可能的宏观分布,把它们各自对应的微观状态数统加起来,便得到该系统所有可到达的微观状态数目,设其为 $\Omega$。根据上述等概率原理,就知道每一微观状态出现的概率是 $1/\Omega$。那么,该系统第 $i$ 种宏观分布出现的概率等于 $W_i/\Omega$。而不同宏观分布出现的概率之间有所差别,关键在于 $W_i$ 有大有小,所以特别地把 $W_i$ 称为第 $i$ 种宏观分布的热力学概率。可见,热力学概率指的是第 $i$ 种宏观分布所对应的微观状态数,它与一般意义下的概率不同,一般的概率值不能大于1,但 $W_i$ 却远远大于1。

**3. 宏观量**

表示系统宏观性质的物理量称为宏观量。而用来描述微观状态粒子性质的量叫作微观量,例如分子的位置、分子间相互作用势能和分子热运动速度、能量等。观测一个宏观量时,常要在规定的某些宏观条件下进行,这就给可能出现的宏观分布及相应的微观配容以一定的约束。不过,一般地,满足约束条件的各种微观状态其数目仍有很多,而且由于实际存在着的一些我们无法完全控制的对系统的微小扰动,使得系统可以在这些可能存在的微观状态之间随机地相互跃迁。尽管我们对宏观量的观测可以进行得很快,但这宏观上看来所用

的短暂时间间隔，与微观状态发生跃迁所经历的时间相比，却长很多，在物理上称为“各态历经”。所以，宏观量其实是在系统所经历的满足给定宏观条件的各种可能微观状态中，相应微观量的统计平均值。

这里讲的“统计平均”，原本是指在一段进行实验观测的微观长时间内，相应的微观量对时间平均的结果。当系统处于平衡态时，宏观量应不随时间改变，即不管观测时间的长短，上述对时间的平均应当一样，因此时间因素对平衡态的统计理论并不重要，关键是要在观测过程中“各态历经”。这样，我们可以换一个角度来处理此统计平均。设想有巨大数目的与原来系统具有相同力学性质、处于相同宏观条件的系统，在同一时刻，这些系统各以一定的概率分别处在原来系统所有可能出现的微观状态上。统计物理把这种作为原来系统不同化身的假想系统之集合叫作“统计系统”。根据对同一时刻系统内各系统状态的观测来求其微观量的统计平均，与前述对一段时间的平均是完全等效的。这好比多次扔一只骰子与同时扔极多只相同的骰子所表现出的统计规律是完全一样的。

**4. 热动平衡态**

在一定的约束条件下，系统可以出现多种分布，每种分布对应一种可能的宏观状态，而实际观察到的宏观状态，乃是各种可能出现的宏观态的平均效果。当系统处于平衡态时，其各个可能的微观状态出现的概率应当是不随时间变化的。只有如此，上述的平均效果才不会随时间变化，或者说，只有概率分布是与时间无关的，所给出的微观量的统计平均值，才会与时间无关。于是表现出所有可被观测的宏观性质都不随时间变化。

以气体分子在空间位置上的分布为例子来说明平衡态所表现的平均分布。设容积为 $V$ 的容器被假想截面分成 $A$、$B$ 两部分，各自容积为 $V_A$、$V_B$。容器中储有一定量气体，热运动中的每一气体分子均可自由穿行这假想截面而往来于 $A$、$B$ 之间。由于分子运动的无规则性以及空间的均匀性，所以任一分子处于 $A$ 部的概率 $p$ 与 $V_A$ 成正比，而处于 $B$ 部的概率 $q$ 与 $V_B$ 成正比；又因为 $p+q=1$，故

$$p=\frac{V_A}{V},\quad q=\frac{V_B}{V} \tag{13-3-4}$$

若一定量气体总共有 $N$ 个可以用编码识别的分子，其中有 $N_1$ 个位于 $A$ 部，$N_2$ 个位于 $B$ 部，每给定一组$(N_1,N_2)$的具体数值，便对应着分子按位置在 $A$、$B$ 两部分空间中的一种分布，显然共有 $N+1$ 种分布。每种确定分布所包含的配容数为$\dfrac{N!}{N_1!\ N_2!}$。

再根据相容的独立事件的概率乘法定理，在 $A$ 中出现 $N_1$ 个有指定编号的分子、而 $B$ 中出现 $N_2=N-N_1$ 个分子，这样一个特定配容的概率应为：$p^{N_1}q^{N-N_1}$，此概率其实只与 $p,q,N_1,N_2$ 有关，而与 $A$ 中究竟是哪些编号的分子无关。这就是说，在给定分布$(N_1,N_2)$下，所包含的各种配容均以此相同的概率出现。于是，根据不相容事件的概率加法定理式，一种分布出现的概率 $P_N(N_1)$是其所对应的各配容出现的概率之和：

$$P_N(N_1)=\frac{N!}{N_1!\ N_2!}p^{N_1}q^{N-N_1} \tag{13-3-5}$$

按此概率分布，可以计算出处于 $A$ 中的分子数之平均值：$\overline{N}_1=Np$，再将 $p=\dfrac{V_A}{V}$代入式中，即有

$$\overline{N}_1=N\frac{V_A}{V} \tag{13-3-6}$$

于是 $A$ 中的平均分子数密度则是：$\bar{n}_1=\dfrac{\bar{N}_1}{V_A}=\dfrac{N}{V}$。

同样可以计算出 $B$ 中的平均分子数密度 $\bar{n}_2$ 也是 $N/V$。这与观测到的平衡态下分子在整个容器中均匀分布是相符的。

以上讨论的是分子在容器内 $A$、$B$ 两部分中的配容和分布，可以进一步给出分子出现在 $A$ 部或 $B$ 部的概率式，进而可得到一个特定配容出现的概率；当然在讨论平衡态下孤立系统中分子在 $\mu$ 相空间的相格中的微观配容时，是不能直接计算一微观配容出现的概率的，只是根据“等概率原理”知道每一微观配容出现的概率都相等。

在上述的例子中，对于不同的分布，$N_1$ 取不同的值，相应的 $P_N(N_1)$会有所不同。取 $N_1=10$，$p=1/3$，与 $N_1$ 对应的 $P_N(N_1)$之值列于表 13-3-1 中。显然，11 种分布各自对应的概率之间存在着明显的差异，其中 $P_{10}(3)$的值最大。我们称其出现概率最大的那一分布为最概然分布。所以，将 $N_1=3$、$N_2=7$ 作为最概然分布。

**表 13-3-1　$N_1=10$，$p=1/3$ 时的二项式概率分布表**

| $N_1$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $P_{10}(N_1)$ | 0.017 | 0.087 | 0.195 | 0.260 | 0.228 | 0.137 | 0.057 | 0.016 | 0.003 | ～0 | ～0 |

将表中的数据绘成离散型随机变量的概率分布图，以 $N_1/N=10$ 时最概然分布对应的 $P_{10}(3)$图线最高，如图 13-3-1 所示。

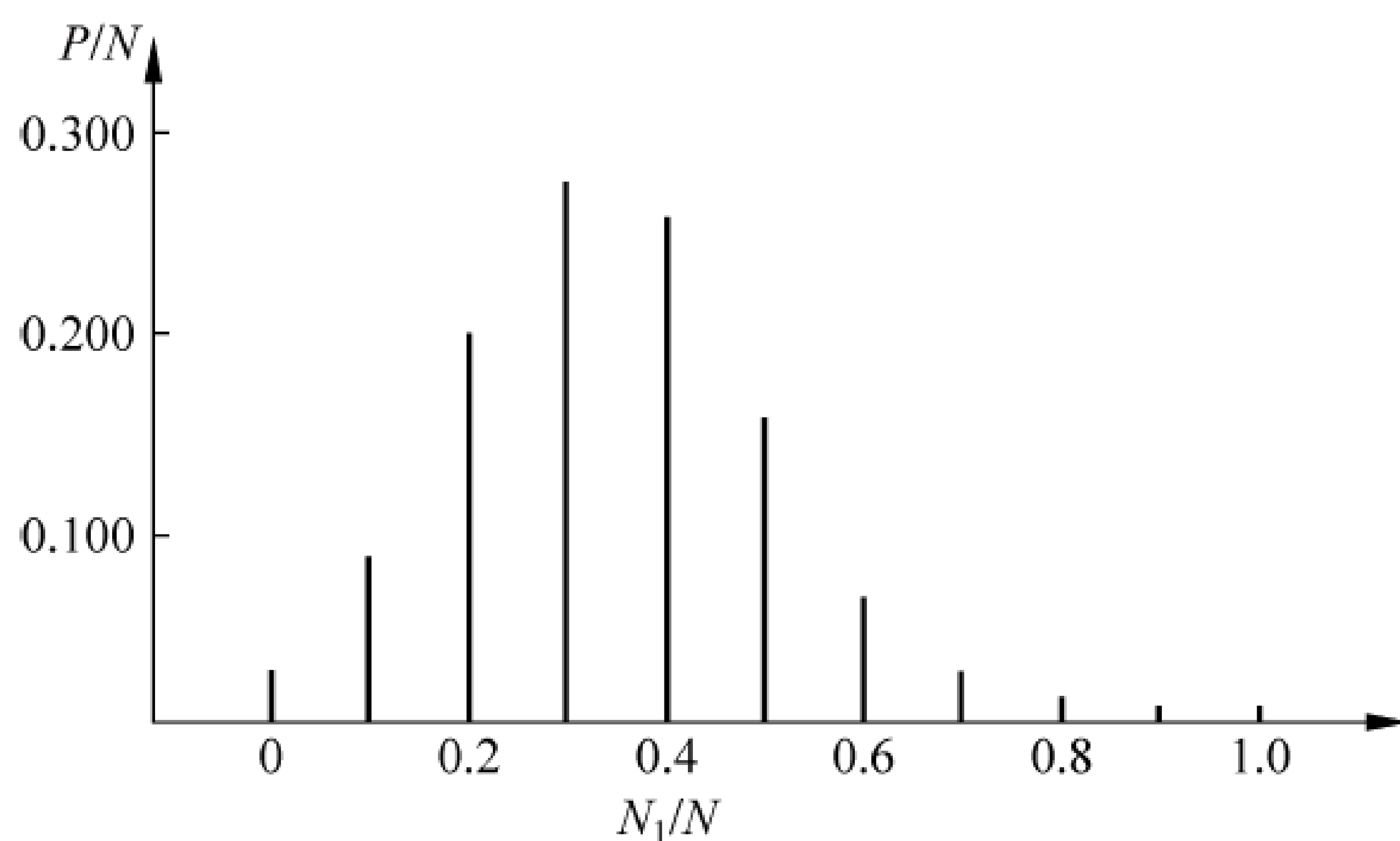



图 13-3-1　$N=10$，$p=1/3$ 时的分布图

当 $N$ 越来越大时，对于 $p=1/3$ 的情况，图线在总体上将于 $N_1/N=1/3$ 附近越来越窄的范围内呈现尖锐的峰，这就意味着最概然分布将趋于平均分布。下面计算出哪一种分布出现的概率最大，并把它与上边求到的平均分布相比较。

要求最概然分布，也就是求 $A$ 部中分子数 $N_1$ 为多大时使得由式(13-3-5)所给出的 $P_N(N_1)$有极大值。通常先取对数，再求极值。

$$\ln P_N(N_1)=\ln N!-\ln N_1!-\ln(N-N_1)!+N_1\ln p+(N-N_1)\ln q$$

由极值条件$\dfrac{\mathrm{d}\ln P_N(N_1)}{\mathrm{d}N_1}=0$，可得：$-\ln N_1+\ln(N-N_1)+\ln p-\ln q=0$。即：$\dfrac{N_1}{N-N_1}=\dfrac{p}{q}$注意到 $p+q=1$，并用式(13-3-4)，可得到：$N_1=Np=N\dfrac{V_A}{V}$。它与平均结果式(13-3-6)一样。

在热力学系统中微观粒子的分布并不像以上讨论的二项分布那么简单，最概然分布与

平均分布是有差异的；但是，最概然分布所对应的配容数远远大于其他分布的配容数，所以，最概然分布与平均分布非常接近。常把给定外界条件下的最概然分布就作为热力学系统平衡态的分布，而其他分布则是对平衡态的偏离，或平衡态附近的涨落。

### 13.3.3　分子运动的统计描述

#### 1. 分子运动方向的统计描述

关于运动方向的统计描述，必须解决两个问题，第一是如何用统计描述表达分子运动的无择优性；第二是如何用统计描述表达在一确定方向附近运动的分子数目。

第一，分子运动无择优方向。

平衡态下，气体宏观性质与方向无关，那么，在微观上，分子的运动必然各向机会均等。为描述这一通性，在空间中任找一点，以其为球心，作任意半径 $r$ 的一个球面，在球面上取确定的一个点 $A$，便给出了由 $O$ 点到 $A$ 点的一个特定方向，如图 13-3-2 所示。某一时刻，气体中每个分子的运动方向都可以用这球面上相应的一个点来代表。我们称这些点为分子运动方向的概率相等，所以，统计地看，这些运动方向的代表点在球面上是均匀分布的。由于分子与容器壁以及分子彼此之间的碰撞，各分子运动方向将发生变化，因此它们的代表点在上述球面上改变位置，但代表点在球面上仍然保持着均匀分布。若共有 $N$ 个分子，那么在此球面上代表点的面密度为 $N/4\pi r^2$。

第二，在一确定方向附近运动的分子数目。

分子运动的方向可以连续变化。也就是，分子运动方向的 $N$ 个分立的代表点各自在球面上的位置是连续变化的，每一分子可能的空间运动方向都有无限多种选择，所以，一个分子运动在某一严格确定方向上的概率等于零。这正如同一分子的速率为一严格确定值的概率为零。

一般地，一气体系统的分子数 $N$ 十分巨大，总有一些分子彼此的运动方向是相近的。例如，在前述球面上取一包括有 $A$ 点的面元 $\Delta S$，运动方向代表点落在 $\Delta S$ 上的那些分子，它们的运动方向就是彼此接近的，如图 13-3-2 所示。这些分子的数目是

$$\Delta N_A = N \times \frac{\Delta S}{4\pi r^2}$$

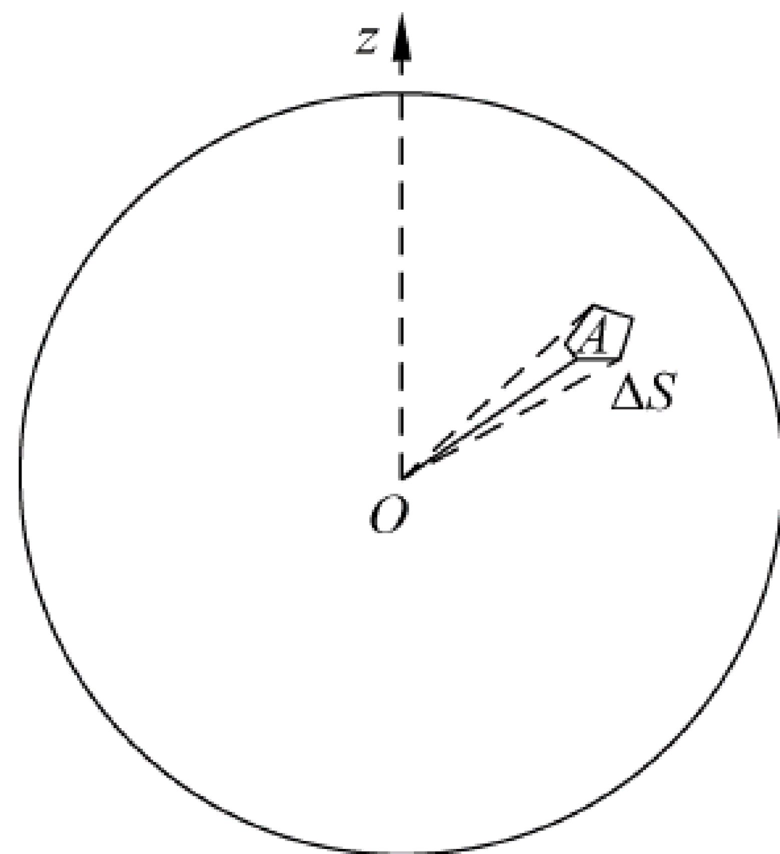


图 13-3-2　分子运动方向的描述

如果由 $\Delta S$ 边缘上各点向球心 $O$ 引直线段，则构成一锥面；一个锥面所围成的空间部分称为立体角，我们记其为 $\Delta\Omega$。立体角是这样度量的：以该锥面的顶点为球心，作任意半径 $r$ 的球面，它与锥面相截所得面积 $\Delta S$，显然 $\Delta S$ 正比于 $r^2$。对于该锥面来说，$\Delta S/r^2$ 是一个常数，就作为此锥面所围立体角的量度，写作 $\Delta\Omega=\dfrac{\Delta S}{r^2}$，其度量单位称为球面度。显然，一个球面对球心所张的立体角就是 $4\pi$ 球面度。于是，$\Delta N_A = N\dfrac{\Delta\Omega}{4\pi}$。

采用球坐标，可以对 $OA$ 方向、$\Delta S$ 面元和立体角 $\Delta\Omega$ 进行更准确的表示。球面上 $A$ 点的位置由矢径 $\overrightarrow{OA}$ 的长度 $r$、$\overrightarrow{OA}$ 与 $z$ 轴的夹角 $\theta$（极角）、$\overrightarrow{OA}$ 在 $xy$ 平面上的投影与 $x$ 轴的夹角 $\varphi$（方位角）来确定。若球面上面元 $\mathrm{d}S$ 对应的极角介于 $\theta\sim\theta+\mathrm{d}\theta$、方位角介于 $\varphi\sim\varphi+$

$\mathrm{d}\varphi$，则 $\mathrm{d}S=r^2\sin\theta\mathrm{d}\varphi$。$\mathrm{d}S$ 对 $O$ 点所张立体角微分元为 $\mathrm{d}\Omega_{\theta,\varphi}=\dfrac{\mathrm{d}S}{r^2}=\sin\theta\mathrm{d}\theta\mathrm{d}\varphi$，运动方向局限在该立体角微分元内的分子数为

$$\mathrm{d}N_{\theta,\varphi}=N\frac{\mathrm{d}\Omega_{\theta,\varphi}}{4\pi}=N\frac{\sin\theta\mathrm{d}\theta\mathrm{d}\varphi}{4\pi} \tag{13-3-7}$$

所以，一分子运动方向局限在该立体角微分元内的概率是$\dfrac{\sin\theta\mathrm{d}\theta\mathrm{d}\varphi}{4\pi}$。

**2. 分子按速度分布及按速率分布的统计描述**

(1) 建立速度空间

所谓的速度空间就是以 $v_x$、$v_y$、$v_z$ 为坐标变量的直角坐标系所描述的一个假想空间。速度矢量可以给出瞬时速度的大小和方向。把所有分子的速度矢量的起始点都平移到上述坐标系的原点，那么，相应速度的大小及方向均由这矢量箭头端点的坐标所完全确定。因此，这些矢量的端点就可以作为分子速度的代表点。于是，分子按速度的分布问题就化为求这些代表点在速度空间的分布问题。当然，速度空间也可以采用球极坐标，如图 13-3-3 所示，一速度矢量的矢径之长，就是其速率 $v$，而极角 $\theta$ 和方位角 $\varphi$ 可以表示出速度的方向。运动方向在 $\theta\sim\theta+\mathrm{d}\theta$、$\varphi\sim\varphi+\mathrm{d}\varphi$ 范围内的分子速度矢量必定落在立体角 $\mathrm{d}\Omega_{\theta,\varphi}=\sin\theta\mathrm{d}\theta\mathrm{d}\varphi$ 之内。

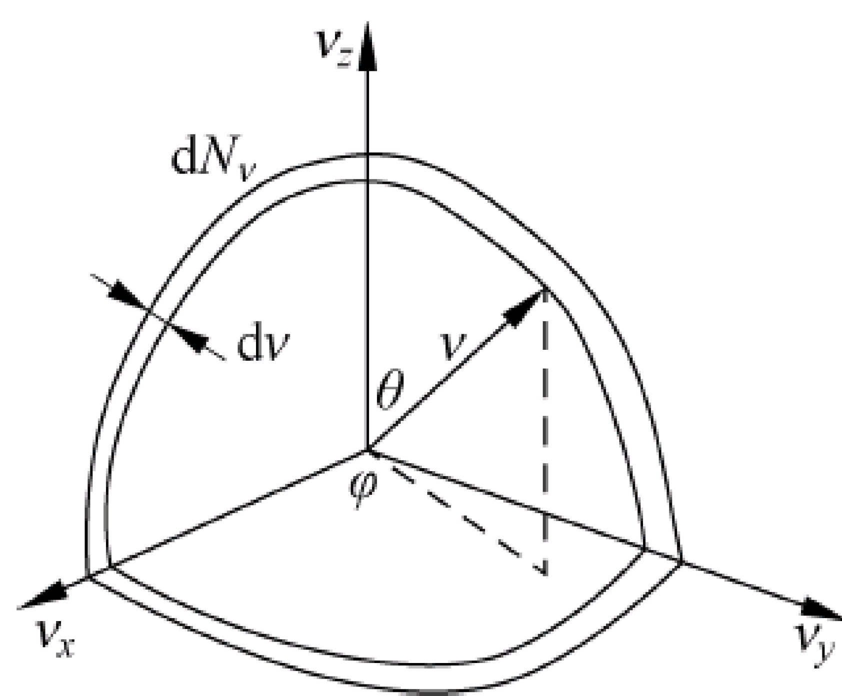


图 13-3-3 速度空间中的球极坐标

(2) 麦克斯韦速度分布律和速率分布律

设系统总分子数为 $N$，速度分量 $v_x$ 在区间 $v_x\sim v_x+\mathrm{d}v_x$ 中的分子数为 $\mathrm{d}N_x$，则分子在这个区间出现的概率为$\dfrac{\mathrm{d}N_x}{N_x}$，用速度分量分布函数 $g_x(v_x)$表示分子在 $v_x$ 附近单位区间出现的概率，则$\dfrac{\mathrm{d}N_x}{N_x}=g_x(v_x)\mathrm{d}v_x$。

系统处于平衡态，各处分子数密度 $n$ 相同，分子向任何方向运动的概率相同，所以对速度分量 $v_y$ 和 $v_z$ 亦有$\dfrac{\mathrm{d}N_y}{N_y}=g_y(v_y)\mathrm{d}v_y$ 和$\dfrac{\mathrm{d}N_z}{N_z}=g_z(v_z)\mathrm{d}v_z$，且分布函数 $g_y(v_y)$和 $g_z(v_z)$与 $g_x(v_x)$的形式相同，故把速度分量分布函数统一记为 $g$。

由于 $v_x$、$v_y$、$v_z$ 互相垂直，互相独立，故我们有理由假设分子沿 $v_x$、$v_y$、$v_z$ 方向的分布互相没有影响，即上述三个概率是相互独立的，所以分子出现在区间($v_x\sim v_x+\mathrm{d}v_x$，$v_y\sim v_y+\mathrm{d}v_y$，$v_z\sim v_z+\mathrm{d}v_z$)内的概率为

$$\frac{\mathrm{d}N_v}{N}=g(v_x)g(v_y)g(v_z)\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z=g_v\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z \tag{12-3-8}$$

其中 $g_v=g(v_x)g(v_y)g(v_z)$即为速度分布函数。

因为分子向任何方向运动的概率相同，所以速度分布应与速度的方向无关，速度分布函数 $g_v$ 只应是速率 $v=\sqrt{v_x^2+v_y^2+v_z^2}$ 的函数，即

$$g_v=g(v_x)g(v_y)g(v_z)=g_v(v_x^2+v_y^2+v_z^2)$$

为满足上式，我们推测 $g(v_x)$应有 $g_v=C\mathrm{e}^{Av_x^2}$($C$、$A$ 为常量)的形式，所以

$$g_v = C\mathrm{e}^{Av_x^2} C\mathrm{e}^{Av_y^2} C\mathrm{e}^{Av_z^2} = C^3 \mathrm{e}^{Av^2}$$

由于 $v$ 很大时分子出现的概率很小，故 $A$ 应取负值，设 $A = -\dfrac{1}{\alpha^2}$，于是

$$\frac{\mathrm{d}N_v}{N} = C^3 \mathrm{e}^{-(v_x^2+v_y^2+v_z^2)/\alpha^2} \mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_z$$

根据归一化条件，即分子出现在所有速度可能值的概率为1，有

$$\int \frac{\mathrm{d}N_v}{N} = C^3 \int_{-\infty}^{\infty} \mathrm{e}^{Av_x^2} \mathrm{d}v_x \int_{-\infty}^{\infty} \mathrm{e}^{Av_y^2} \mathrm{d}v_y \int_{-\infty}^{\infty} \mathrm{e}^{Av_z^2} \mathrm{d}v_z = 1$$

求得，$C = \left(\dfrac{1}{\pi\alpha^2}\right)^{1/2}$。于是速度分布函数和速度分量分布函数分别为

$$g_v = \left(\frac{1}{\pi\alpha^2}\right)^{3/2} \mathrm{e}^{-(v_x^2+v_y^2+v_z^2)/\alpha^2} \tag{13-3-9}$$

$$g(v_x) = \left(\frac{1}{\pi\alpha^2}\right)^{1/2} \mathrm{e}^{-v_x^2/\alpha^2} \tag{13-3-10}$$

由于速率分布与速度方向无关，只考虑分子出现在速率区间 $v \sim v + \mathrm{d}v$ 在速度空间对应图中的球壳，其体积为 $4\pi v^2 \mathrm{d}v$。把速度分布律 $\dfrac{\mathrm{d}N_v}{N} = \left(\dfrac{1}{\pi\alpha^2}\right)^{3/2} \mathrm{e}^{-(v_x^2+v_y^2+v_z^2)/\alpha^2} \mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_z$ 中的 $v_x^2 + v_y^2 + v_z^2$ 记为 $v^2$，由于 $v$ 在球壳内有相同的取值，所以只要把 $\mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_z$ 换为 $4\pi v^2 \mathrm{d}v$，即得到分子出现在速率空间 $v \sim v + \mathrm{d}v$ 内的概率为

$$\frac{\mathrm{d}N}{N} = 4\pi \left(\frac{1}{\pi\alpha^2}\right)^{3/2} \mathrm{e}^{-v^2/\alpha^2} v^2 \mathrm{d}v$$

这就是速率分布律。由上式求出分子速率平方的统计平均值为 $\overline{v}^2 = \dfrac{3}{2}\alpha^2$，与 $\overline{E}_\mathrm{k} = \dfrac{1}{2}m\overline{v}^2 = \dfrac{3}{2}kT$ 比较可知 $\alpha^2 = \dfrac{2kT}{m}$。

于是得到麦克斯韦速度分布律为

$$\frac{\mathrm{d}N_v}{N} = \left(\frac{m}{2\pi kT}\right)^{3/2} \mathrm{e}^{-m(v_x^2+v_y^2+v_z^2)/2kT} \mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_z \tag{13-3-11}$$

麦克斯韦速度分量分布律为

$$\frac{\mathrm{d}N_x}{N} = \left(\frac{m}{2\pi kT}\right)^{1/2} \mathrm{e}^{-mv_x^2/2kT} \mathrm{d}v_x \tag{13-3-12}$$

麦克斯韦速率分布律为

$$\frac{\mathrm{d}N}{N} = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \mathrm{e}^{-mv^2/2kT} v^2 \mathrm{d}v \tag{13-3-13}$$

### 13.3.4　理想气体压强公式的推导比较及讨论

**1. 单个分子冲力统计法**

设理想气体盛于边长为 $a$、$b$、$c$ 的长方体容器中处于平衡态，如图13-3-4所示。每个分子质量为 $m$，共有 $N$ 个分子。设第 $i$ 个分子的速度为 $\boldsymbol{v}_i$，其直角坐标分量分别为 $v_{ix}$、$v_{iy}$、$v_{iz}$，且满足 $v_i^2 = v_{ix}^2 + v_{iy}^2 + v_{iz}^2$。根据理想气体模型，分子之间及分子与器壁间的碰撞是完

全弹性的,则第 $i$ 个分子与器壁 $A_1=bc$ 面碰撞后,其速度分量变为 $-v_{ix}$、$v_{iy}$、$v_{iz}$。同理与 $A_2$ 面碰撞后,其速度分量变为 $v_{ix}$、$-v_{iy}$、$v_{iz}$,与 $A_3$ 碰撞后速度分量为 $v_{ix}$、$v_{iy}$、$-v_{iz}$。则在 $x$ 方向上,分子 $i$ 获得器壁施予的冲量为 $f_i\Delta t_i=\Delta p_i=p_{i2}-p_{i1}=m(-v_{ix})-mv_{ix}=-2mv_{ix}$,式中 $f_i$ 是器壁施于分子 $i$ 的冲力。$\Delta t_i$ 是分子 $i$ 与器壁碰撞的时间,由牛顿第三定律,该分子施予器壁的冲力 $f'_i$ 的大小亦为 $f_i$,与 $f_i$ 的方向相反,且有 $f'_i\Delta t_i=-f_i\Delta t_i=2mv_{ix}$。

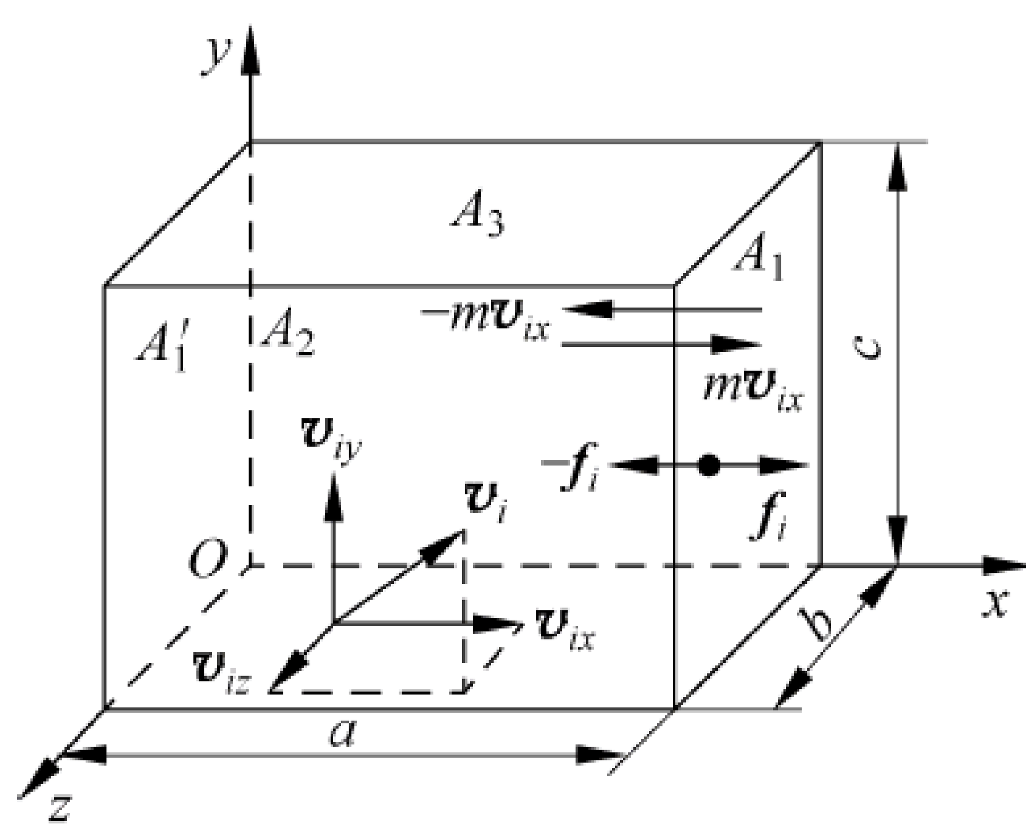



图 13-3-4 分子速度的分解

如果忽略分子 $i$ 在面 $A_1$ 和 $A'_1$ 之间运动时与其他分子的碰撞,则分子 $i$ 连续两次与 $A_1$ 的碰撞时间为 $\dfrac{2a}{v_{ix}}$,则单位时间内,该分子与面 $A_1$ 碰撞次数为 $\dfrac{v_{ix}}{2a}$,分子 $i$ 施予器壁的总冲量为 $I=$碰撞次数$\times f'_i\Delta t_i$,即

$$I=\frac{v_{ix}}{2a}\times 2mv_{ix}=m\frac{v_{ix}^2}{a} \tag{13-3-14}$$

另一方面,设 $i$ 分子在单位时间内施予器壁的平均冲力为 $\bar{F}_I$,则

$$I=\int_0^l f'_i\,\mathrm{d}t=\bar{F}_i(1-0)=\bar{F}_i \tag{13-3-15}$$

由式(13-3-14)和式(13-3-15)得:$\bar{F}_i=m\dfrac{v_{ix}^2}{a}$。因单个分子施予 $A_1$ 的冲力是间断的,而容器内大量分子不断地与面 $A_1$ 碰撞,则使 $A_1$ 受到一个持续的作用力,考虑大数目 $N$ 个分子,则

$$\begin{aligned}\bar{F}&=\bar{F}_1+\bar{F}_2+\cdots+\bar{F}_N=\frac{mv_{1x}^2}{a}+\frac{mv_{2x}^2}{a}+\cdots+\frac{mv_{Nx}^2}{a}\\&=\frac{m}{a}\sum_{i=1}^{N}v_{ix}^2=\frac{Nm}{a}\sum_{i=1}^{N}\frac{v_{ix}^2}{N}\end{aligned} \tag{13-3-16}$$

根据统计平均假定有

$$\sum_{I=1}^{N}\frac{v_{ix}^2}{N}=\bar{v}_x^2=\frac{1}{3}\bar{v}^2\quad(\bar{v}_x^2=\bar{v}_y^2=\bar{v}_z^2) \tag{13-3-17}$$

则,式(13-3-16)变为

$$\bar{F}=\frac{1}{3}\frac{mN}{a}\bar{v}^2 \tag{13-3-18}$$

由压强的定义得

$$P=\frac{\bar{F}}{S}=\frac{mN}{3abc}\bar{v}^2=\frac{1}{3}nm\bar{v}^2=\frac{1}{3}\rho\bar{v}^2=\frac{2}{3}n\bar{w} \tag{13-3-19}$$

式中 $n=\dfrac{N}{v}=\dfrac{N}{abc}$,$\rho=nm$ 为气体密度,$\bar{w}=\dfrac{1}{2}m\bar{v}^2$ 为分子的平均平动能。

**2. 总冲量统计平均法**

如上所述,单位时间内,单个分子 $i$ 施予器壁面 $A_1$ 的总冲量为 $2mv_{ix}\dfrac{v_{ix}}{2a}=m\dfrac{v_{ix}^2}{a}$,则

$N$ 个分子在单位时间内施予 $A_1$ 的总冲量为

$$I=\sum_{i=1}^{N}m\frac{v_{ix}^2}{a}=\frac{m}{a}\sum_{i=1}^{N}v_{ix}^2 \tag{13-3-20}$$

设容器内分子数目 $N$ 巨大，则器壁 $A_1$ 受到的力可以看成是一个连续均匀的平均力 $\bar{F}$，则由 $I$ 的含义和冲量定理知，在平衡态下，在任意时间段内，应有

$$I=\bar{F} \tag{13-3-21}$$

式(13-3-21)表明，单位时间内 $A_1$ 面所受分子的总冲量在量值上等于 $A_1$ 所受的连续均匀的平均力 $\bar{F}$，因此由式(13-3-17)和式(12-3-21)得

$$\bar{P}=\frac{\bar{F}}{bc}=\frac{I}{bc}=\frac{m}{abc}\sum_{I=1}^{N}v_{ix}^2=\frac{Nm}{V}\sum\frac{v_{ix}^2}{N}=\frac{1}{3}nm\bar{v}^2=\frac{1}{3}\rho\bar{v}^2=\frac{2}{3}n\bar{w} \tag{13-3-22}$$

式中有关量含义与式(13-3-19)相同。

**3. 微分冲量元法**

取任意形状的容器盛有处于平衡态的理想气体，每个分子的质量为 $m$，单位体积内分子数为 $n=\dfrac{N}{V}$，按麦克斯韦速度分布律，此时分子具有各种可能的速度，将其分成若干组，假定每组内分子具有大小相等方向一致的速度，每组内分子数分别为 $n_1$、$n_2$、…、$n_i$，则

$$n=n_1+n_2+\cdots+n_k=\sum_{i=1}^{k}n_i\quad(k<N,k\in N) \tag{13-3-23}$$

如图 13-3-5 所示，在垂直于 $x$ 轴的器壁上任取一面积微元 $\mathrm{d}A$，由以上讨论知，分子 $i$ 在一次碰撞中施予 $\mathrm{d}A$ 的冲量为 $2mv_{ix}$，全部速度为 $\bar{v}_i$ 的第 $i$ 组分子 $n_i$ 中，与 $\mathrm{d}A$ 在 $\mathrm{d}t$ 相碰的只是位于体积元内 $\mathrm{d}Av_{it}\mathrm{d}t$ 的分子数 $\mathrm{d}N_i$，即 $\mathrm{d}N_i=n_iv_{ix}\mathrm{d}A\mathrm{d}t$，对应的冲量元 $\mathrm{d}I_i=2mv_{ix}\mathrm{d}N_i=2n_imv_{ix}^2\mathrm{d}A\mathrm{d}t$。实际上，在体积元内，只有 $v_{ix}>0$ 的分子才与 $\mathrm{d}A$ 碰撞，当气体处于平衡态时，从统计平均上讲，$v_{ix}>0$ 和 $v_{ix}<0$ 的分子各占一半，故有

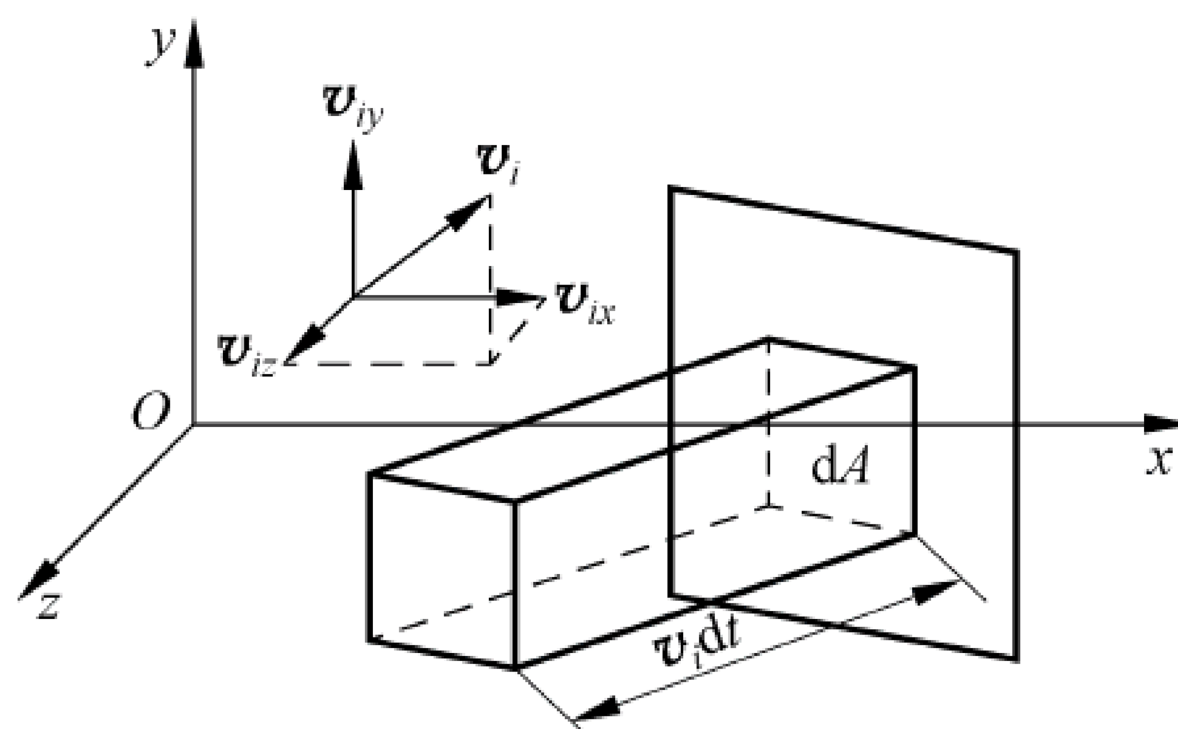


图 13-3-5 分子一次碰撞中施予 $\mathrm{d}A$ 的冲量

$$\mathrm{d}I=\frac{1}{2}\mathrm{d}I_i=\frac{1}{2}\sum_{i=1}^{k}2n_imv_{ix}^2\mathrm{d}A\mathrm{d}t=\sum_{i=1}^{k}n_imv_{ix}^2\mathrm{d}A\mathrm{d}t \tag{13-3-24}$$

因 $\mathrm{d}I$ 是持续作用在 $\mathrm{d}A$ 上，则 $\mathrm{d}I$ 与 $\mathrm{d}t$ 之比即为气体施予器壁 $\mathrm{d}A$ 上的持续作用力，因此得

$$P=\frac{\bar{F}}{S}=\frac{\mathrm{d}I/\mathrm{d}t}{\mathrm{d}A}=\frac{\mathrm{d}I}{\mathrm{d}t}\frac{1}{\mathrm{d}A}=\sum_{i=1}^{k}n_imv_{ix}^2 \tag{13-3-25}$$

由统计平均值的计算式和式(12-3-23)得

$$\bar{v}_i^2=\frac{\sum\limits_{i=1}^{k}n_iv_{ix}^2}{\sum\limits_{i=1}^{k}n_i}=\frac{\sum\limits_{i=1}^{k}n_iv_{ix}^2}{n} \tag{13-3-26}$$

将式(13-3-26)代入式(13-3-25)并考虑式(13-3-17)得

$$P=nm\bar{v}_x^2=\frac{1}{3}nm\bar{v}^2=\frac{1}{3}\rho\bar{v}^2=\frac{2}{3}n\bar{w} \tag{13-3-27}$$

式中各量与式(13-3-19)含义相同。

**4. 积分平均求和法**

设在容器壁 $O$ 点附近有一单位立方体,其内有 $n$ 个理想气体分子,其中具有速度 $v_i$ 者为 $n_i$ 个。以 $n_{iqg}$ 表示在方向间隔 $\theta \to \theta t\,\mathrm{d}\theta, \varphi \to \varphi t\,\mathrm{d}\varphi$ 之间的分子数,如图 13-3-6 所示。根据分子在各个方向上运动机会均等的统计原则,$n_{iqg}$ 与 $n_i$ 之比应等于方向间隔 $\mathrm{d}\theta\mathrm{d}\varphi$ 所对应的面积元 $\mathrm{d}s = R\,\mathrm{d}\theta R\sin\theta\mathrm{d}\varphi$ 与相应球面之比,即

$$\frac{n_{iQ\varphi}}{n_i} = \frac{R^2\sin\theta\mathrm{d}\theta\mathrm{d}\varphi}{4\pi R^2} = \frac{1}{2}\,\frac{1}{2\pi}\sin\theta\mathrm{d}\theta\mathrm{d}\varphi$$

单位立方体在 $\mathrm{d}t$ 时间内,在方向间隔 $\mathrm{d}\theta\mathrm{d}\varphi$ 中,到达器壁 $O$ 点处 $\mathrm{d}A$ 面积元的分子数应包含在图 13-3-7 所示的斜柱体内,此斜体的中轴线与器壁 $O$ 点的法线成 $\theta$ 角,斜柱体高为 $v_i\mathrm{d}t\cos\theta$,则斜柱体积为 $v_i\cos\theta\mathrm{d}A\,\mathrm{d}t$,其内分子总数为 $\mathrm{d}N_i$,即

$$\mathrm{d}N_i = n_{iq\varphi}v_i\cos\theta\mathrm{d}A\,\mathrm{d}t = \frac{1}{4\pi}n_i v_i\cos\theta\sin\theta\mathrm{d}\theta\mathrm{d}\varphi\mathrm{d}A\,\mathrm{d}t \tag{13-3-28}$$

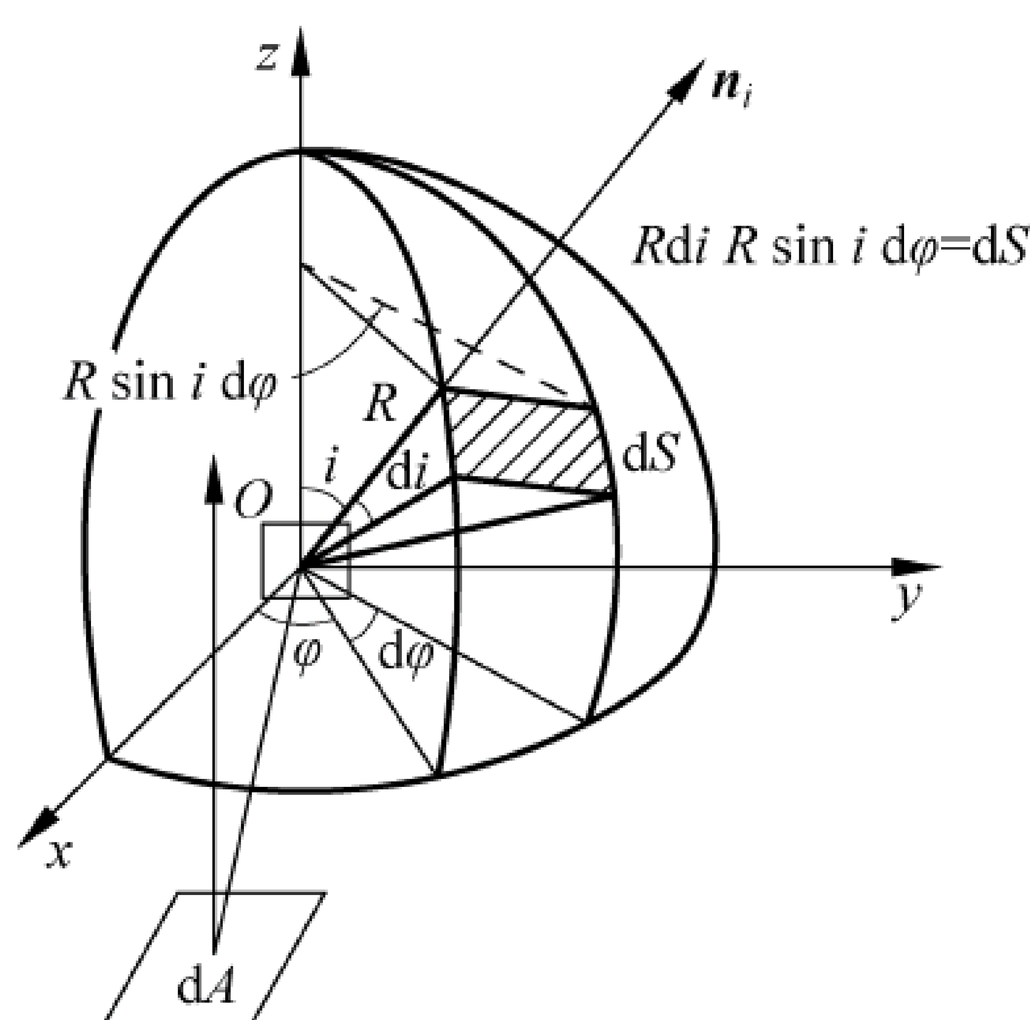



图 13-3-6 立方体角内理想气体分子

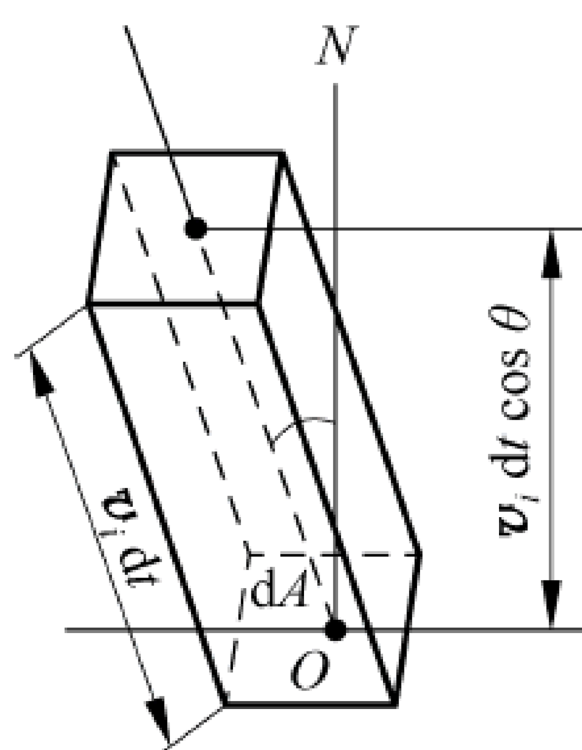



图 13-3-7 单位立方体的体积

一个分子与器壁每碰撞一次施予其冲量为 $2mv_i\cos\theta$($\theta$ 为入射角,假设其碰撞为对称的完全弹性散射),因此,在 $\mathrm{d}t$ 时间内,在方向间隔内的分子,施予器壁 $O$ 点 $\mathrm{d}A$ 处的冲量元为

$$\mathrm{d}I_i = 2mv_i\cos\theta\mathrm{d}N_i = \frac{1}{2\pi}n_i m v_i^2\cos^2\theta\sin\theta\mathrm{d}\theta\mathrm{d}\varphi\mathrm{d}A\,\mathrm{d}t \tag{13-3-29}$$

则容器内各个方向具有速率 $v_i$ 的分子在 $O$ 点产生的平均压强为

$$\begin{aligned}\bar{P}_i &= \int \frac{\mathrm{d}I_i}{\mathrm{d}t}\,\frac{1}{\mathrm{d}A} \\ &= n_i m v_i^2\int_0^{-\frac{\pi}{2}}\cos^2\theta\sin\theta\mathrm{d}\theta\int_0^{2\pi}\frac{1}{2\pi}\mathrm{d}\varphi \\ &= \frac{1}{3}n_i m v_i^2 \end{aligned} \tag{13-3-30}$$

按麦克斯韦速度分布律,式(13-3-29)中 $v_i$ 取各种可能值,则每个速度值对应的平均压强为

$$P_1 = \frac{1}{3}n_1 m v_1^2,\quad P_2 = \frac{1}{3}n_2 m v_2^2, \cdots, P_k = \frac{1}{3}n_k m v_k^2$$

当容器内分子数足够大时,则上述平均压强之和即为宏观的稳定压强,即

$$P=P_1+P_2+\cdots+P_k=\frac{1}{3}m(n_1v_1^2+n_2v_2^2+n_3v_3^2+\cdots+n_kv_k^2)=\frac{1}{3}m\sum_{i=1}^{k}n_iv_i^2$$

由式(13-3-26)得

$$P=\frac{1}{3}nm\overline{v}^2=\frac{1}{3}\rho\overline{v}^2=\frac{2}{3}n\overline{w} \tag{13-3-31}$$

式中各量与式(13-3-19)含义相同。

**5. 讨论**

从以上过程可以看出,理想气体压强公式的推证体现了分子物理学的研究方法,是统计力学的基础之一,主要在以下几个方面。

(1) 在分子的运动速度与光速相比小得多的低速范围内,经典力学原理在分子运动层次仍成立,推证过程应用了动量定理和牛顿第三定律。

(2) 单个分子的运动服从力学规律,大量分子的运动服从统计规律,平衡态是条件。

首先,在理想气体模型中,压强公式 $P=\frac{1}{3}nm\overline{v}^2=\frac{1}{3}\rho\overline{v}^2=\frac{2}{3}n\overline{w}$,反映了 $\overline{v}^2$ 或 $\overline{w}$、$n$ 或 $\rho$ 与 $P$ 这三组统计平均之间的关系,是统计规律而不是力学规律,因此,对单个分子谈压强是没有意义的。

其次,按统计规律,任何统计平均值都有涨落存在,涨落反映了测量值与统计平均之间的偏差,因此,压强公式中的速度方均涨落、平均平动能涨落、粒子数密度涨落或质量密度涨落都会存在,因而,就其时间效应来说,时间微元应是宏观足够小、微观足够大,以致有大量分子与器壁发生碰撞,才能忽略短暂不持续的碰撞涨落,才会有稳定的气体压强。

(3) 分子与器壁的碰撞,实质是气体分子受器壁分子的散射,而且入射角不一定等于反射角,但从统计观点出发。某个分子入射角为 $\theta_1$,反射角为 $\theta_2$ 时,必有入射角 $\theta_2$,反射角为 $\theta_1$ 的另一分子,因为分子是全同的。因此对大量分子的统计平均,必有入射角等于反射角。

(4) 在分子与器壁连续两次碰撞之间,忽略了分子之间的碰撞,这是因为从统计观点,气体处于平衡态时,若有的分子因碰撞而改变了自己的速度,必有其他分子因与之碰撞也改变了速度。另一方面,在平衡态下,从气体分子速度分布与方向等统计概念出发并运用对大量分子的分组取和及求统计平均等统计方法导出了式(13-3-17)、式(13-3-21)、式(13-3-23)、式(13-3-24)和式(13-3-26),它们都是压强公式的直接依据,也就是说,这些式子都反映了气体压强产生的物理本质,亦反映了宏观力学规律与经典统计规律的联系。

## 13.4 光学中的矢量方法

在力学里,我们知道力、位移、速度、加速度、动量等是矢量,矢量的合成满足矢量的叠加原理(几何方法就是平行四边形定则或三角形定则)。比如,力的合成满足力的叠加原理:如果在一质点上同时作用几个力(力的强度不是太大),则这些力产生的效果等同于一个合力总的效果,而不互相影响,不同力的这种叠加不同于数学的代数和(为什么?)。另一方面我们还知道,力的线性叠加原理是以力的独立作用原理为前提的,这些力各自产生自己的效果而不互相影响的。应当指出,非线性科学研究表明,当力的强度足够大时,力的叠加不满足线性叠加原理,也就是说,当矢量的强度足够大时,矢量的叠加不满足线性叠加原理,也不再遵从独立作用原理。

光波是电磁波，其电场强度是矢量，因此，在光强不太大的条件下，光波电场强度的叠加也满足线性叠加原理。当光强很强时，比如高功率激光束射向某种介质，在其内光波电场强度的叠加可能不再满足线性叠加原理，这就是非线性光学的研究领域。

### 13.4.1 线性叠加原理与波的独立传播

光波的线性叠加原理可以表述为：空间某位置几列光波相遇，在该位置所引起的振动矢量等于各列波分别在该位置所引起的振动矢量的矢量和。该原理可认为是物理学中广泛的矢量叠加原理在光学中的具体应用。

观察和研究表明：几列光波可以保持各自的特点(比如其频率、波长、振幅和振动方向等)同时通过同一介质，好像在各自的传播过程中没有遇到其他波一样，也就是说从几个振源发出的波相遇于同一区域时，只要振动不十分强烈，就可以各自保持自己的特性，彼此不受影响，这就是波动独立性的表现。在相遇区域内，介质质点的合位移是各波分别单独传播时在该点所引起的位移的矢量和。因此，可以简单地把各波的分位移按照矢量加法叠加起来，这就是波的叠加性，这种叠加性是以独立性为条件的。光波电场强度矢量叠加的数学表达式为

$$\boldsymbol{E}(p,t)=\boldsymbol{E}_1(p,t)+\boldsymbol{E}_2(p,t)+\cdots$$

通常情况下，波动方程是线性微分方程，简谐波的表达式就是它的一个解，如果有两个独立的函数都满足同一个给定的微分方程，那么这两个函数的和也必然是这个微分方程的解，这就是两个独立的波的叠加的数学意义。

光波的叠加原理来源于波的独立传播定律，它们在一定的条件下才成立，这条件是介质和光波的强度。光在其中服从叠加原理的介质，称为“线性介质”，线性介质的电磁特性和光波的光矢量大小无关。这就要求光场不能太强。当光强很大的时候，如激光束，波的叠加原理就不能成立。不同介质当中使波的叠加原理不能成立的光强度一般不相同，它与介质本身的特性有关。

由发光过程的物理机制可知，一个普通光源包含着数目极多的随机状态的原子辐射体，每一个激发态原子大概只在 $10^{-9}$ 秒的时间间隔内持续发射一列偏振波，以后又将会有新的偏振波列不时地发出来，辐射是有间歇的，无规则的，每一瞬时大量原子发射的和不同瞬时同一原子发射的电磁波，在相位、偏振态等特征方面均无固定联系，理论上可以是一种完全随机的分布。如果我们使某一列偏振光波分列开来，并分别使它们经过不同的光程后在空间某一区域重新会合，当各列光程差对应的时间差不超过一次辐射所持续的时间时，我们就会看到相位、偏振态有固定联系的波列的叠加，事实上这种叠加便是干涉(时间相干性)和偏振的共同基础。

### 13.4.2 相干叠加和非相干叠加

为了分别研究相位、偏振态有固定关系的两种叠加现象，我们定义：各分场波矢量相位之间有固定联系的叠加为相干叠加，反之，为非相干叠加。我们认为任何一个可观察的相干叠加和非相干叠加现象都是瞬时相干叠加的时间统计平均效果。因为一次发光只能持续 $10^{-8}$ 秒，没有记录仪可以观察到光在 $10^{-8}$ 秒时间内的瞬时相干叠加现象。

由此可见，在几乎同一直线上的，同频率的两个电磁振动叠加时，必须区别两种情况：

(1) 两振动的相位差始终保持不变，合振动平均强度可以大于也可以小于分振动强度之和。在这种情况下就可能在较长时间内观察到干涉现象，通常称作频率相同、振动方向几

乎平行并且在观察期内相位差保持不变的两个振动是相干的。其实只要两个振动方向上互相不垂直即可，取它们同向的分量叠加，与此垂直方向分量的叠加暂时略去。这两振动的叠加就是相干叠加，也就是出现了可见的稳定的干涉花样。

（2）两振动的相位差在观察时间内无规则地改变，合振动的平均强度简单地等于分振动强度之和，不出现干涉花样，这种振动叠加为不相干叠加。

现代光学分为相干光学和非相干光学两大类，而相干光学中又有两种研究方法：标量波理论和矢量波理论，本节只讨论矢量波的叠加。当各分场矢量方向彼此平行时，矢量问题转化为标量问题来处理。光学中的干涉现象是指电场矢量彼此平行的两列或两列以上的光波的叠加，这种叠加的时间平均结果是产生稳定的、强弱相间的能量分布，形成干涉条纹。类似的分析还可以用于处理衍射问题（请读者参阅本书第 9 章关于光的衍射的详细讨论）。至于光的各种偏振态的形成则是与电场矢量彼此正交的同频率波列的相干叠加相联系的，这种叠加的时间平均结果产生各种稳定的偏振光。

一般情况下的光学现象中，参与叠加的电场矢量方向既可以不平行也可以不正交；既有相干叠加又有非相干叠加，因而构成了各种部分相干光和部分偏振光的光学现象，它们都是同一物理过程——叠加过程中的不同侧面反映。

### 13.4.3 衍射现象是相干叠加的结果

衍射问题存在特殊性，它在理论和方法上近似可以用波的叠加原理来解决问题。

光波在传播过程中如果遇到障碍物而使光波受到限制时，就会发生衍射现象，在障碍物后形成背离光线光学（直线传播）的衍射花样。从理论上说衍射问题应在由障碍物提供的某种边界条件下求麦克斯韦方程组而得。但是，一般情况下求严格的矢量解非常困难，所以，采用惠更斯子波假设基础上的惠更斯-菲涅耳原理作为研究衍射现象的理论基础。这个原理对衍射过程最本质的近似是认为衍射场是由透过障碍物的入射光波波振面上的假设（假想）子波在障碍物后相干叠加的结果。由于障碍物是一种近场作用，因此在离障碍物足够远处，惠更斯-菲涅尔定理能够给出令人满意的解答。这种近似处理的物理本质就是把满足相干条件的子波光源发出的光波在障碍物处透过的波阵面上的场矢量与子波的波源联系起来了。可见，叠加原理同样也是解决衍射的基本原理。区别在于衍射现象中处理的是同一个波阵面上无数个连续分布的假想子波发出的同振动方向的波列的相干叠加问题。

### 13.4.4 光波矢量叠加的数学方法

#### 1. 三角函数法

设同频率、振动方向相同的两列单色光波：$\boldsymbol{E}_1(p,t)=\boldsymbol{E}_{01}(p)\cos(\omega t-\varphi_1(p))$和$\boldsymbol{E}_2(p,t)=\boldsymbol{E}_{02}(p)\cos(\omega t-\varphi_2(p))$在空间交叠，如图 13-4-1 所示。$\boldsymbol{E}_1,\boldsymbol{E}_2$ 是空间任一点 $P$ 的光振动，式中的符号小写 $p$ 表示空间位置 $P$ 点与光源距离，根据光波叠加原理，$P$ 点的合成光振动的振幅为

$$\begin{aligned}\boldsymbol{E}(p,t)&=\boldsymbol{E}_1(p,t)+\boldsymbol{E}_2(p,t)\\&=\boldsymbol{E}_{01}(p)\cos(\omega t-\varphi_1(p))+\boldsymbol{E}_{02}(p)\cos(\omega t-\varphi_2(p))\end{aligned}$$

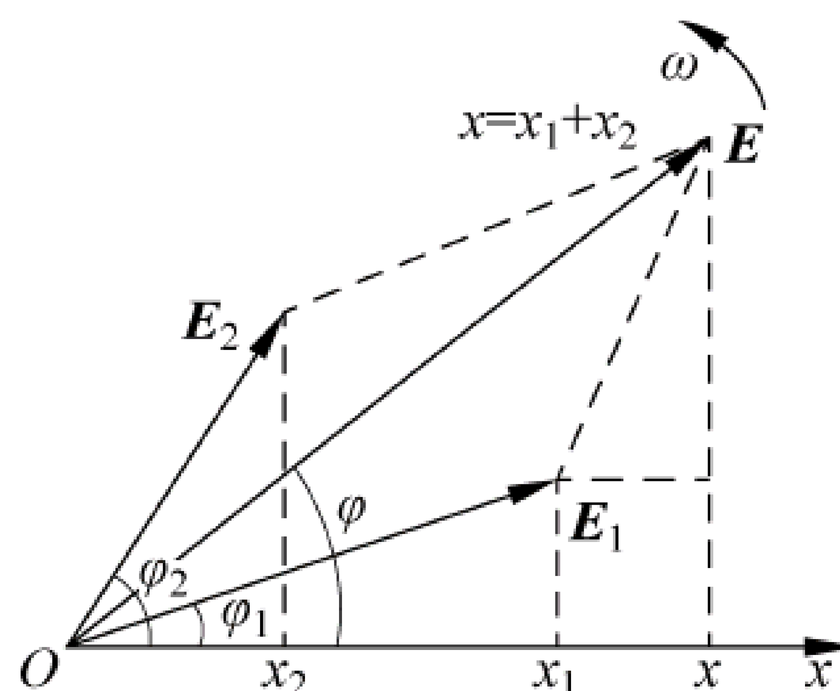


图 13-4-1 两列波相干叠加

直接利用三角函数关系式可求得合振动强度为

$$E_0^2(p)=E_{01}^2(p)+E_{02}^2(p)+2\boldsymbol{E}_{01}(p)\cdot\boldsymbol{E}_{02}(p)\cos(\varphi_2-\varphi_1)\tag{13-4-1}$$

$$\tan\varphi=\frac{E_{01}\sin\varphi_1+E_{02}\sin\varphi_2}{E_{01}\cos\varphi_1+E_{02}\cos\varphi_2} \tag{13-4-2}$$

可见，合成光振动仍是相同频率的单色光波，但振幅和相位与组成它的成分不相同。

因为振动的强度正比于振幅的平方，从式(13-4-1)可以看出，在相位差 $\varphi_2-\varphi_1$ 为任意角的情况下(但不能随时间变化，为什么?)，两个振动叠加时，合振动的强度不等于两个分振动之和。若有多列光波相干叠加，使用这种方法也比较麻烦。

**2. 复振幅法**

若有 $N$ 个振动方向，频率相同的单色光波在空间相遇，那么它们之中第 $i$ 列光波可用下式来描述：$\widetilde{E}(p,t)=E_{0i}(p)\mathrm{e}^{\varphi_i(p)}\cdot\mathrm{e}^{-\mathrm{i}\omega t}=E_{0i}\mathrm{e}^{-\mathrm{i}\omega t}$。

当满足波的叠加原理时，$P$ 点总的光振动为

$$\widetilde{E}(p,t)=\sum_{i=1}^{N}\widetilde{E}_i(p,t)=\mathrm{e}^{-\mathrm{i}\omega t}\left[\sum_{i=1}^{N}\widetilde{E}_{0i}(p)\right]$$

在只有两列波相干叠加时，应当首先写出每列光波的复振幅，然后相加，即 $\widetilde{E}=\widetilde{E}_1+\widetilde{E}_2$ 利用 $\cos a=\dfrac{\mathrm{e}^{\mathrm{i}a}+\mathrm{e}^{-\mathrm{i}a}}{2}$可以得到合振动。

$P$ 点的复振幅为

$$\widetilde{E}(p)=\left[\sum_{i=1}^{N}\widetilde{E}_{0i}(p)\right]=\widetilde{E}_{01}(P)\mathrm{e}^{\mathrm{i}\varphi_1}+\widetilde{E}_{02}(P)\mathrm{e}^{\mathrm{i}\varphi_2}$$

$P$ 点的强度为

$$E_0^2(P)=\widetilde{E}(p)\cdot\widetilde{E}^*(p)=[\widetilde{E}_{01}(P)\mathrm{e}^{\mathrm{i}\varphi_1}+\widetilde{E}_{02}(P)\mathrm{e}^{\mathrm{i}\varphi_2}][\widetilde{E}_{01}(P)\mathrm{e}^{\mathrm{i}\varphi_1}+\widetilde{E}_{02}(P)\mathrm{e}^{\mathrm{i}\varphi_2}]^*$$

$$=E_{01}^2(p)+E_{02}^2(p)+2\widetilde{E}_{01}(p)\cdot\widetilde{E}_{02}(p)\cos(\varphi_2-\varphi_1) \tag{13-4-3}$$

再求合振动的初相位 $\varphi(p)$，把式右边的负数形式开展为三角函数

$$E_0(p)\mathrm{e}^{\mathrm{i}\varphi(p)}=E_{01}(p)\cos\varphi_1+E_{02}(p)\cos\varphi_2+\mathrm{i}[E_{01}\sin\varphi_1+E_{02}\sin\varphi_2]$$

根据复数的性质，得到

$$\tan\varphi=\frac{E_{01}\sin\varphi_1+E_{02}\sin\varphi_2}{E_{01}\cos\varphi_1+E_{02}\cos\varphi_2} \tag{13-4-4}$$

**3. 矢量合成法**

设同频率、振动方向相同的两列单色光波：

$$\begin{aligned}\boldsymbol{E}_1(p,t)&=\boldsymbol{E}_{01}(p)\cos(\omega t-\varphi_1(p))\\ \boldsymbol{E}_2(p,t)&=\boldsymbol{E}_{02}(p)\cos(\omega t-\varphi_2(p))\end{aligned} \tag{13-4-5}$$

在空间交叠。根据振动数学分析理论，我们可以用两个旋转矢量 $\boldsymbol{E}_1$、$\boldsymbol{E}_2$ 来描述它们，这些矢量被称为振幅矢量。它们的长度代表光振动的振幅大小，它们与某一给定轴($x$ 轴)的夹角等于所表示光振动的初相位。若振幅矢量绕 $O$ 点以角速度 $\omega$ 逆时针方向旋转，则两矢量的末端在 $Ox$ 轴上投影的运动正好与式(13-4-5)相符合，这也是一个振幅矢量可以和一个简谐振动对应的原因。随着 $\boldsymbol{E}_1$、$\boldsymbol{E}_2$ 的同步旋转，显然它们的合矢量 $\boldsymbol{E}$(利用图 13-4-1 平行四边形加法求出)也绕 $O$ 点以同一角速度逆时针旋转，且 $\boldsymbol{E}$ 的末端在 $Ox$ 轴上的投影的运动也为一简谐振动：

$$\boldsymbol{E}(p,t)=\boldsymbol{E}_0(p)\cos(\omega t-\varphi(p))$$

因为 $\boldsymbol{E}$ 在 $Ox$ 轴上的投影总是等于 $\boldsymbol{E}_1$、$\boldsymbol{E}_2$ 在 $Ox$ 轴上投影之和，即 $\boldsymbol{E}_1=\boldsymbol{E}_1+\boldsymbol{E}_2$，所以两个

单色光波在某一点的光振动叠加可以通过它们的振幅矢量相加来求得，由图 13-4-1 可见：

$$E_0^2(p)=E_{01}^2(p)+E_{02}^2(p)+2\boldsymbol{E}_{01}(p)\cdot\boldsymbol{E}_{02}(p)\cos(\varphi_2-\varphi_1)$$

$$\tan\varphi=\frac{E_{01}\sin\varphi_1+E_{02}\sin\varphi_2}{E_{01}\cos\varphi_1+E_{02}\cos\varphi_2}$$

矢量图对于求多个谐振动的合成特别有用，这是应用矢量多边形的加法求出矢量 $\boldsymbol{E}_1$、$\boldsymbol{E}_2$、…、$\boldsymbol{E}_n$ 的合矢量 $\boldsymbol{E}$，图 13-4-2 表示了多个振幅矢量的多边形加法，两相邻矢量间的夹角为两相应振动的相位差。这个方法是分析多光束干涉的方法之一。

图 13-4-2　多两列波相干叠加

**4. 正交分解法**

场矢量的正交分解法是一个非常有用的分析方法，它不仅揭示了合场矢量由分场矢量叠加而得，而且还应包括：当我们在任一选定的正交坐标系中对某一场矢量作分解时，该矢量可表示为对应正交分量的线性组合。

用正交分解法解决光波叠加问题时，一般遵循以下步骤。

（1）建立正交坐标系；

（2）对已知各场矢量做出正交分解，并用叠加原理求出各个分量：

$$\boldsymbol{E}_x=\sum\boldsymbol{E}_{ix},\quad i=1,2,\cdots$$

$$\boldsymbol{E}_y=\sum\boldsymbol{E}_{iy},\quad i=1,2,\cdots$$

$$\boldsymbol{E}_z=\sum\boldsymbol{E}_{iz},\quad i=1,2,\cdots$$

（3）由步骤(2)给出的结果，作进一步计算，解决各种具体问题。

例如，普通光源发射的自然光的叠加特征。

选用笛卡儿坐标系，如图 13-4-3 所示，研究对象是沿着 $z$ 轴正方向传播的普通光源射出的某种那个频率的自然光。由于电磁波的横波性质，在与 $z$ 轴垂直的平面上，将自然光中的任一偏振光进行正交分解，则其电场矢量在任意选定的 $x$，$y$ 正交轴上的瞬时分量如图 13-4-4 所示，自然光的大量原子发射的电磁波场矢量在任意选定的 $x$，$y$ 正交轴上的瞬时分量如图 13-4-5 所示，并由下式给出：

$$\begin{aligned}\boldsymbol{E}_x&=\boldsymbol{E}_{1x}+\boldsymbol{E}_{2x}+\boldsymbol{E}_{3x}+\cdots+\boldsymbol{E}_{nx}\\\boldsymbol{E}_y&=\boldsymbol{E}_{1y}+\boldsymbol{E}_{2y}+\boldsymbol{E}_{3y}+\cdots+\boldsymbol{E}_{ny}\end{aligned}\qquad(13\text{-}4\text{-}6)$$

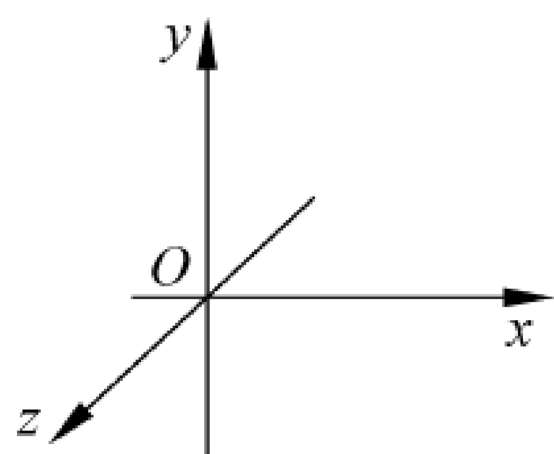


图 13-4-3　笛卡儿坐标系

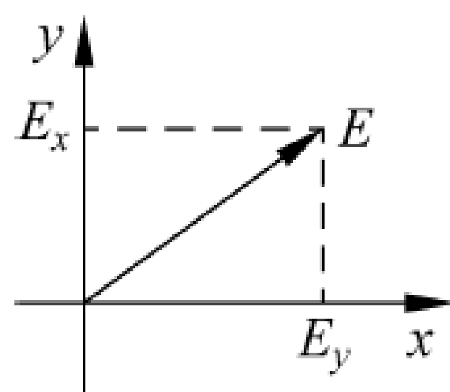


图 13-4-4　矢量的正交分解

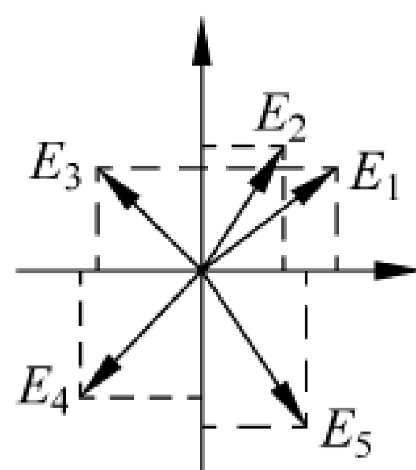


图 13-4-5　自然光矢量的正交分解

总合成场矢量为

$$\boldsymbol{E}=\boldsymbol{E}_x+\boldsymbol{E}_y\qquad(13\text{-}4\text{-}7)$$

在传播过程中各处的光强为

$$I=\langle\boldsymbol{E}\cdot\boldsymbol{E}^*\rangle\qquad(13\text{-}4\text{-}8)$$

将式(13-4-6)、式(13-4-7)代入式(13-4-8)得到

$$I=\langle \boldsymbol{E}\cdot\boldsymbol{E}^{*}\rangle$$
$$=\sum|\boldsymbol{E}_{ix}|^2+\sum|\boldsymbol{E}_{iy}|^2+\sum_{i\neq j}\boldsymbol{E}_{ix}\cdot\boldsymbol{E}_{jx}^{*}+\sum_{i\neq j}\boldsymbol{E}_{iy}\cdot\boldsymbol{E}_{jy}^{*}+\sum_{\substack{i=1\\j=1}}^{n}\boldsymbol{E}_{ix}\cdot\boldsymbol{E}_{jy}^{*}+\sum_{\substack{i=1\\j=1}}^{n}\boldsymbol{E}_{iy}\cdot\boldsymbol{E}_{jx}^{*} \tag{13-4-9}$$

根据普通光源发光过程的随机性质,式(13-4-9)中:

$$\left\langle\sum_{i\neq j}^{n}\boldsymbol{E}_{ix}\cdot\boldsymbol{E}_{jx}^{*}\right\rangle\equiv 0$$
$$\left\langle\sum_{i\neq j}^{n}\boldsymbol{E}_{iy}\cdot\boldsymbol{E}_{jy}^{*}\right\rangle\equiv 0 \tag{13-4-10}$$

式(13-4-10)分别反映相位随机变化和偏振态空间分布的统计特征。

根据矢量代数理论可知:

$$\left\langle\sum\boldsymbol{E}_{ix}\cdot\boldsymbol{E}_{iy}^{*}\right\rangle=\left\langle\sum\boldsymbol{E}_{iy}\cdot\boldsymbol{E}_{ix}^{*}\right\rangle=0 \tag{13-4-11}$$

它反映任意两个正交矢量点积永远为零的事实。

将式(13-4-9)、式(13-4-10)和式(13-4-11)结果代入式(13-4-8),则得到 $I=\left\langle\sum_{1}^{n}|\boldsymbol{E}_{ix}|^2+\sum_{1}^{n}|\boldsymbol{E}_{iy}|^2\right\rangle=I_x+I_y$,并且有

$$I_x=\left\langle\sum_{1}^{n}|\boldsymbol{E}_{ix}|^2\right\rangle=I_y=\left\langle\sum_{1}^{n}|\boldsymbol{E}_{iy}|^2\right\rangle=\frac{1}{2}I \tag{13-4-12}$$

式(13-4-12)说明,自然光对于任意选定的正交坐标系,总能分解为两个不相干的、正交的偏振光,且光强相等,振幅均为$\sqrt{1/2}$。可见,用正交分解法进行一定的计算,不仅很容易得到普通光学中有关自然光的分析结论,而且加深了对自然光的理解,因为上述计算过程本身说明自然光是大量原子随机辐射叠加的结果。如果从瞬时态和平均效果角度分析,在理论上自然光的本质还可以理解为无数随机变化的瞬时椭圆偏振光的非相干叠加。

### 13.4.5 光波矢量叠加法的例子

**1. 菲涅耳单缝衍射的振幅矢量作图法**

现在,我们用另一种方法——振幅矢量作图法来定性分析菲涅耳衍射。如果我们将第一个半周期带再分成 $N$ 个小带,小带的分法为:以 $P$ 点为球心,分别以

$$r_0+\frac{1}{N}\frac{\lambda}{2},r_0+\frac{2}{N}\frac{\lambda}{2},r_0+\frac{3}{N}\frac{\lambda}{2},\cdots,r_0+\frac{N}{N}\frac{\lambda}{2}$$

为半径作球,在第一个半周期带上分成 $N$ 个子带,每一个子带在 $P$ 点产生一个振动,而 $N$ 个振动的合成必定是 $E$。第一个子带发出的次波到达 $P$ 点的振动,其振幅矢量用 $\Delta A_1$ 表示,第二个子带到达 $P$ 点的振动,其振幅矢量用 $\Delta A_2$ 表示,$\Delta A_2$ 的大小比 $\Delta A_1$ 小一些(因为倾斜因子收缩每一个子带的振幅),位相要比 $\Delta A_1$ 落后$\frac{\pi}{N}$$\left(\text{光程差为}\frac{\lambda}{2}\right.$,相当于 π 的位相差,现光程差为$\frac{1}{N}\frac{\lambda}{2}$,故位相差为$\left.\frac{\pi}{N}\right)$。第三、第四……子带可依次类推。根据惠更斯-菲涅

耳原理，第一个半周期带的次波到达 $P$ 点的振动应该是这些带的振动合成，其相应的振幅矢量也应该是这些带所产生的振动的振幅矢量的合成图。图 13-4-6(a)就表示 $N=17$ 个子带振幅矢量的合成图。如果把这个波带分成无限多($N\to\infty$)的小部分，这时折线图就变成与半圆形相差很小的一个圆弧，如图 13-4-6(b)所示。同时圆弧在 $M$ 点的方向与 $O$ 点的方向相反，因为这个波带最后一个圆部分在 $P$ 点产生的振动位相显然与开始第一个子带所产生的相位相反，因为两者到 $P$ 点的距离相差 $\frac{\lambda}{2}$，$OM_1$ 就是第一个波带在 $P$ 点引起的合振动的复振幅。

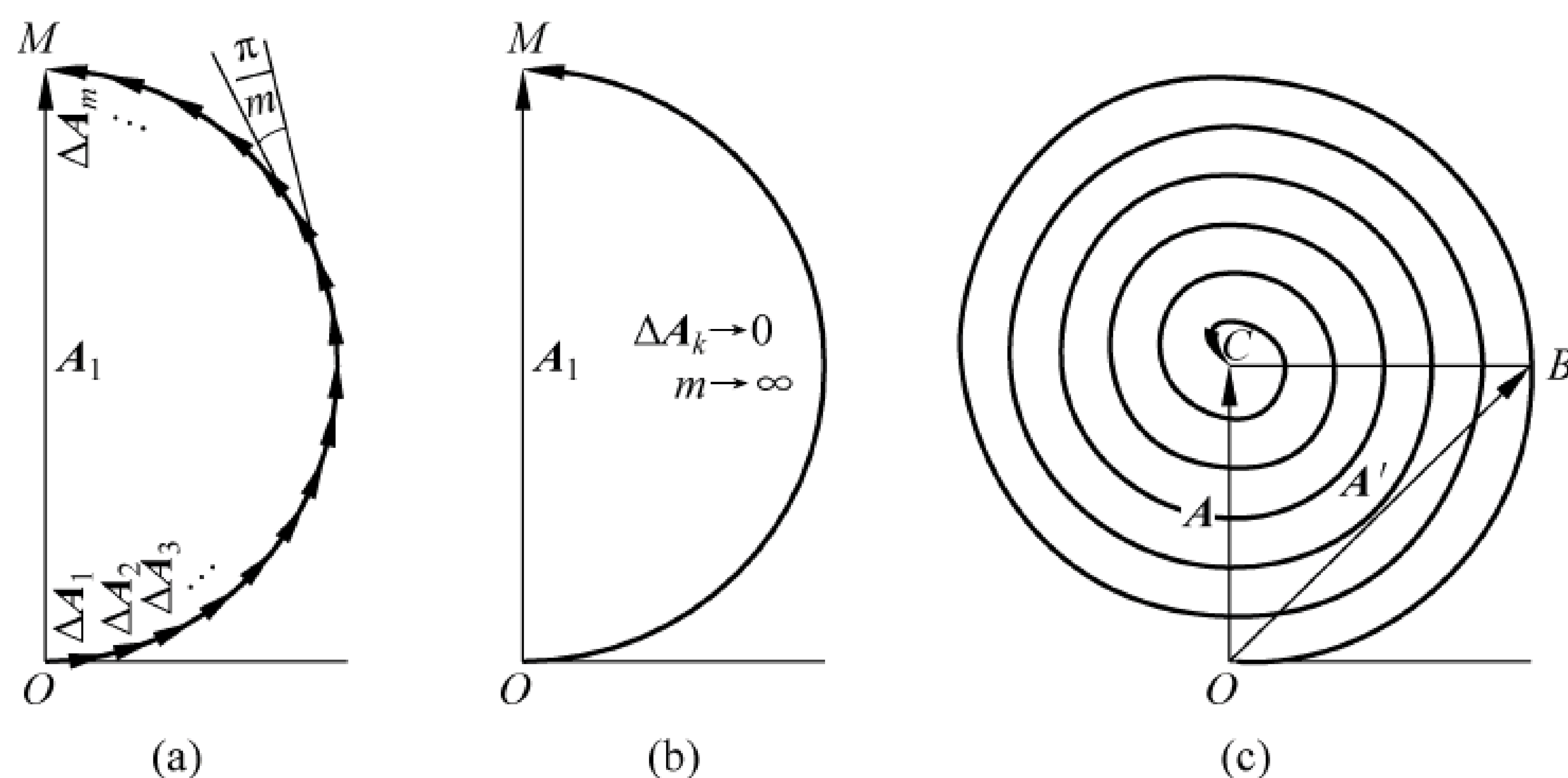


图 13-4-6　半波带振幅的矢量合成方法——三角形(平行四边形)法

第二个波带在 $P$ 点的合振动又是怎么样的呢？根据上面的讨论，就在第一波带的振幅矢量图中以 $M_1$ 为起点再画一个稍微小一些，和半圆稍许偏差的曲线，如图 13-4-6(c)所示。波带数愈多，则其振幅矢量的合成图就愈趋近一螺线。从图 13-4-6(c)中还可以看出当奇数带时，矢量的末端 $M_m$ 点出现在曲线的上面，合振动振幅 $|E_P|=OM_m=\frac{1}{2}OM_1+\frac{1}{2}OM_m$，即，$|E_P|=\frac{1}{2}|E_1|+\frac{1}{2}|E_m|$；同样，当 $m$ 是偶数，$M_m$ 点在图下半部，$P$ 点合振动的振幅为 $|E_P|=\frac{1}{2}|E_1|+\frac{1}{2}|E_m|$，如图 13-4-7 所示。

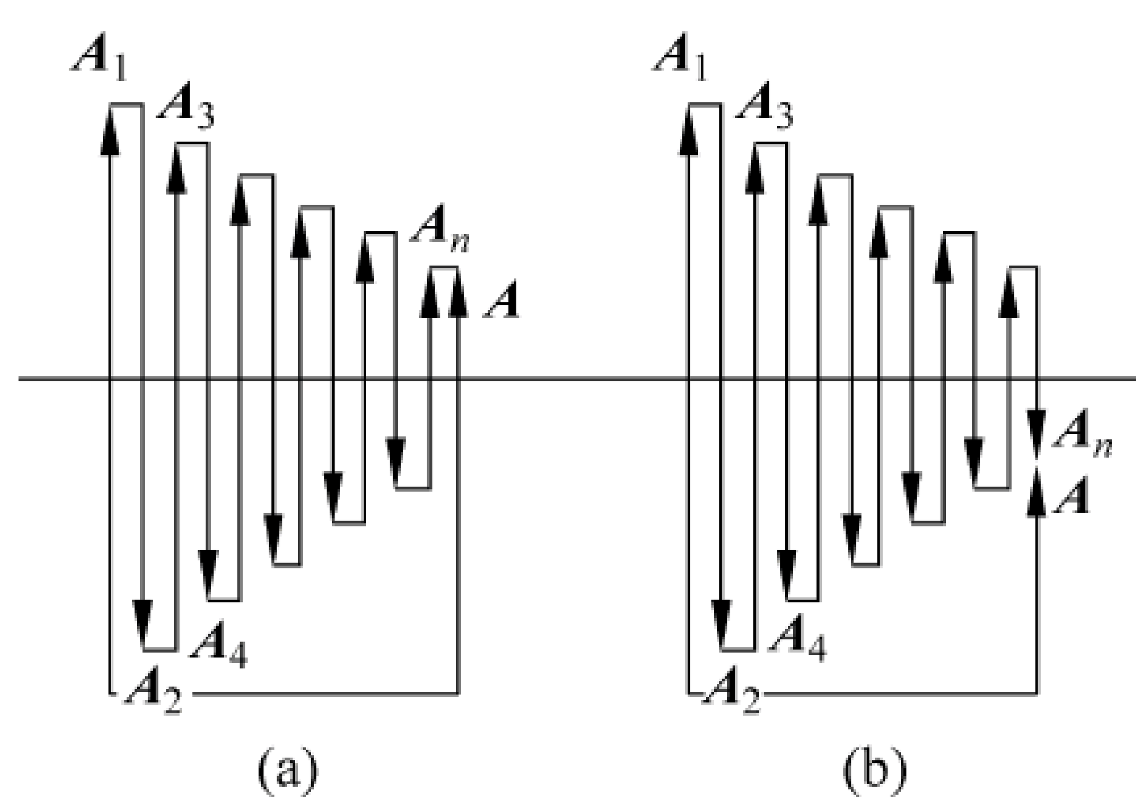


图 13-4-7　半波带法的振幅矢量合成图
(a) $n$ 为奇数；(b) $n$ 为偶数

在自由传播情况下，这螺旋线就一直绕到半径趋向于零，最后到达圆心 $C$。由 $O$ 到 $C$ 作合成矢量，其长度即为整个光波自由传播时在 $P$ 点产生的振幅。如图 13-4-6(c)所示。

根据同样原理，利用振幅矢量图可以求出单色光经任意形状的衍射屏在屏后 $P$ 点所产生的合振动振幅。此时应注意合振动振幅不仅由此屏对 $P$ 点所露出的波带的数目决定，还与每个波带露出的面积有关。如有一个半周期带有一半面积被衍射屏挡住，则此半周期带所产生的振幅矢量应缩小一半。

**2．单缝衍射光强分布公式的振幅矢量作图法**

单缝强度分布公式可用振幅矢量作图法推得。假设单缝平面的波阵面(即露出的波面)分成很多的等分，例如九等分，因为每一等分宽度相等，它们对屏上某一点的距离相差很小，

所以每一等分在屏上 $P$ 点所引起的振幅 $\alpha$ 相等。至于振动的位相,则对每个等分来说是不同的,但两相邻等分在屏上 $P$ 点所引起振动的位相差 $\delta$ 是一定的。则振幅矢量合成如图 13-4-8 所示,$A_0$ 代表此时的合振幅。如果把波阵面分成无限多等分,则矢量图为一圆弧,其长仍为 $A_0$,且合振幅为 $A_\theta$,等于这圆弧的弦长,从图中可以看出,这段弧对圆心所张的角应为 $2\beta$,因为首末两矢量的位相差就是 $2\beta$。设弧的半径为 $R$,并从圆心作 $A_\theta$ 的垂线,则由几何关系,可以得到

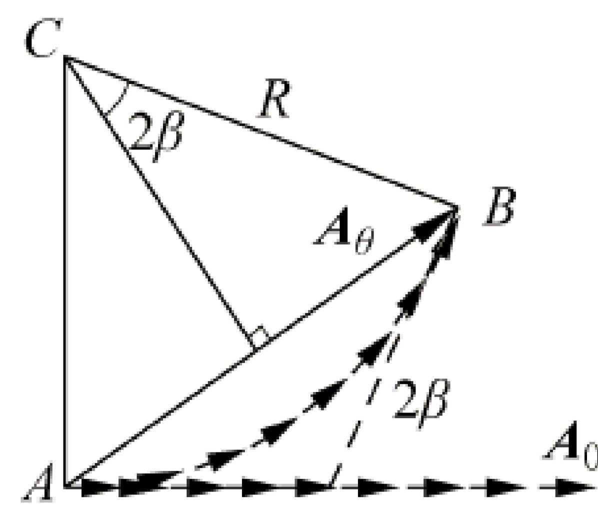



图 13-4-8　光强分布公式的振幅矢量作图法

$$\sin\beta=\frac{\frac{1}{2}A_\theta}{R},\quad A_\theta=2R\sin\beta$$

因而有:$\dfrac{A_\theta}{A_0}=\dfrac{\text{弦}}{\text{弧}}=\dfrac{2q\sin\beta}{R2\beta}=\dfrac{\sin\beta}{\beta}$,即,$A_\theta=A_0\dfrac{\sin\beta}{\beta}$,最后有

$$I=I_0\frac{\sin^2\beta}{\beta^2}$$

## 13.5　自然态场势与数学物理方程

许多自然现象的数学形式都可以是偏微分方程式,特别是很多重要的物理力学及工程过程的基本规律的数学描述都是偏微分方程,例如流体力学、电磁学的基本定律都是如此。这些反映物理及工程过程的规律的偏微分方程就是所谓的数学物理方程。当然,几何学中的很多问题也是可以用偏微分方程来描述的。

### 13.5.1　数学物理方程

人们对偏微分方程的研究,从微分学产生后不久就开始了。例如,18 世纪初期对弦的横向振动研究,其后对热传导理论的研究,以及对流体力学、对位函数的研究,都获得相应的数学物理方程及其有效的解法。到 19 世纪中叶,进一步从个别方程的深入研究逐渐形成了偏微分的一般理论,如方程的分类、特征理论等,这便是经典的偏微分方程理论的范畴。

然而到了 20 世纪,随着科学技术的不断发展,在科学实践中提出了数学物理方程的新问题,电子计算机的出现为数学物理方程的研究提供了强有力的实现手段。又因为数学的其他分支(如泛函分析、拓扑学、群论、微分几何等)也有了迅速发展,为深入研究偏微分方程提供了有力的工具。

**1. 数学物理方程的基本性质**

20 世纪关于数学物理方程的研究有了前所未有的发展,这些发展呈如下特点和趋势:

其一,在许多自然科学及工程技术中提出的问题的数学描述大多是非线性偏微分方程,即使一些能用线性偏微分方程作近似处理的问题,由于研究的深入,也必须重新考虑非线性效应。对非线性偏微分方程研究,难度大得多,然而线性偏微分方程的已有结果,能提供很多有益的启示。

其二,实践中的过程是由很多因素联合作用和相互影响的。所以其数学模型多是非线性偏微分方程组。如反应扩散方程组、流体力学方程组、电磁流体力学方程组和辐射流体方程组等,在数学上称为双曲-抛物方程组。

其三，数学物理方程不再只是描述物理学、力学等工程过程的数学工具。目前在化学、生物学、医学、农业、环保领域，甚至在经济学等社会科学领域都不断提出一些非常重要的偏微分方程。

其四，一个实际模型的数学描述，除了描述过程的方程(或方程组)外，还应有定解条件(如初始条件及边值条件)。传统的描述，这些条件是线性的，逐点表示的。而现在提出的很多定解条件是非线性的，特别是非局部的。对非局部边值问题的研究是一个新的非常有意义的领域。

其五，与数学其他分支的关系。例如几何学中提出了很多重要的非线性偏微分方程，如极小曲面方程、调和映照方程等。泛函分析、拓扑学及群论等现代工具在偏微分方程的理论研究中被广泛应用。广义函数的应用使得经典的线性微分方程理论更系统完善。再就是计算机的广泛应用，计算方法的快速发展，特别是有限元广泛的应用，使得对偏微分方程的研究得以在实践中实现和检验。

**2. 关于偏微分方程的基本概念和相关知识**

称形如

$$f(x_1,x_2,\cdots,u,u_{x1},u_{x2},\cdots,u_{x_1x_2},\cdots)=0 \tag{13-5-1}$$

的方程为 $u$ 关于独立变量 $x_i$ 的偏微分方程。这里：$x_1,x_2,\cdots$是相互独立的变量，$u=u(x_1,x_2,\cdots)$是独立变量的未知函数；$u_{x_1},u_{x_2},\cdots,u_{x_1x_1},u_{x_1x_2},\cdots$是未知函数 $u$ 关于独立变量 $x_1,x_2,\cdots$的偏导数，$(x_1,x_2,\cdots,x_n)\in D\in \mathbf{R}^n$，$n\geqslant 2$ 这里 $\mathbf{R}^n$ 为 $n$ 维欧几里得空间 $D$——$\mathbf{R}^n$ 中的开区域。

偏微分方程的解：如果存在一个足够光滑的函数 $u=u(x_1,x_2,\cdots,x_n)$，即方程(13-5-1)中出现的 $u=u(x_1,x_2,\cdots,x_n)$及其偏导数在区域 $D$ 上存在，且在 $\overline{D}$ 中连续，在区域 $D$ 上满足方程(13-5-1)的函数 $u=u(x_1,x_2,\cdots,x_n)$称为方程(13-5-1)的解。

比如，可以很简单地证明函数 $u(x,y)=(x+y)^3$，$u(x,y)=\sin(x-y)$都是方程 $u_{xx}-u_{yy}=0$ 的解。其理论基础是偏微分方程解的唯一性定理。

偏微分方程的阶：关于未知函数 $u$ 的偏微分中阶数最高的阶数。比如，$u_{xx}+2uu_{xy}+u_{yy}=\mathrm{e}^y$ 是一个二阶方程，$u_{xxy}+xu_{yy}+8u=7y$ 是一个三阶方程。

线性方程：所有的未知函数和它们的偏导数都是线性的，方程中所有的系数只依赖于独立的变量。

拟线性方程：所有最高阶的偏导数是线性的，但方程是非线性的。

非线性方程：最高阶的偏导数是非线性的。比如，$yu_{xx}+2xyu_{xy}+u_y=x^2$ 是一个二阶的线性方程；$u_xu_{xx}+xuu_y=\cos x$ 是一个二阶拟线性方程；$(u_{xy})^2+5u_x+e^yu=y^2$ 是一个二阶非线性方程，等等。

一般形式：$n$ 个独立变量的二阶线性偏微分方程是

$$\sum_{i,j=1}^{n}A_{ij}u_{x_ix_j}+\sum_{i=1}^{n}B_iu_{x_i}+Fu=G \tag{13-5-2}$$

这里的 $A_{ij}=A_{ji}$，$B_i$，$F$ 和$G$ 是仅仅依赖于$n$ 个独立变量的函数。

齐次方程：如果 $G=0$；非齐次方程：如果 $G\neq 0$。

偏微分方程的解是一般解，不同于普通的微分方程的解得依赖于一些常数。

比如，其一，如果 $u_{xy}=0$，我们可以得到：$u_x(x,y)=f(x)\Rightarrow u(x,y)=g(x)+h(y)$，

这里 $g(x)$ 和 $h(y)$ 是一般的连续可微函数。其二，如果假定 $u=u(x,y,z)$ 和 $u_{yy}=2$，就可以获得一般解：$u(x,y,z)=y^2+yf(x,z)+g(x,z)$。这里的 $f$ 和 $g$ 是一般的连续可微方程，和只有有限的线性独立函数的 $n$ 阶偏微分方程不同的是它可以包含无限个线性独立函数；其三，如果我们对 $u_x+u_y=0$ 使用线性变换：

$$\begin{cases}\xi=x+y\\ \eta=x-y\end{cases}$$

则，可以得到 $2u_\eta=0$ 并获得一般解：

$$u(x,y)=f(x+y)$$

这里 $f(x+y)$ 是一般的连续可微函数，包含无数的函数。例如：$(x+y)^n$，$\sin n(x+y)$，$\cos n(x+y)$，$\exp n(x+y)(n=1,2,\cdots)$，这些函数都是线性无关的。

### 13.5.2 算子

一种把一个函数变换成另一个函数的数学运算法则。

**例** $L[u]=\dfrac{\partial u}{\partial x}+\dfrac{\partial^2 u}{\partial x\partial y}+\dfrac{\partial^3 u}{\partial y^3}$ 和 $M[u]=\dfrac{\partial^2 u}{\partial x^2}-x^2\dfrac{\partial^2 u}{\partial y^2}$。这里 $L=\dfrac{\partial}{\partial x}+\dfrac{\partial^2}{\partial x\partial y}+\dfrac{\partial^3}{\partial y^3}$ 和 $M=\dfrac{\partial^2 u}{\partial x^2}-x^2\dfrac{\partial^2 u}{\partial y^2}$ 称为微分算子。

### 13.5.3 分离变量法

分离变量法的主要思想。其一，将方程中含有各个变量的项分离开来，从而原方程拆分成多个更简单的只含1个自变量的常微分方程；其二，运用线性叠加原理，将非齐次方程拆分成多个齐次的或易于求解的方程；其三，利用微积分知识、级数求解知识以及其他巧妙方法，求出各个方程的通解；其四，运用叠加原理，将这些通解“组装”起来。

分离变量法是求解偏微分方程最基本最常用的方法。主要的理论依据是线性方程的叠加原理和 Sturm-Liouville 理论。最核心的思想是将偏微分方程的求解化为对常微分方程的求解。

下面就有界弦的自由振动的定解问题讨论

$$\begin{cases}\dfrac{\partial^2 u}{\partial t^2}-a^2\dfrac{\partial^2 u}{\partial x^2}=0, & 0<x<l\\ u|_{x=0}=0, u|_{x=l}=0, & t>0\\ u|_{t=0}=\phi(x), \left.\dfrac{\partial u}{\partial t}\right|_{t=0}=\psi(x), & 0<x<l\end{cases}$$

观察注意其特点是：方程齐次，边界齐次。

端点会引起波的反射，弦有限长，波在两端点之间往返反射。两列反向行进的同频率的波形成驻波。驻波的特点：(1) 没有波形的传播，即各点振动相位与位置无关，按同一方式随时间振动，可统一表示为 $T(t)$。(2)各点振幅随位置而变化而与时间无关，用 $X(x)$ 表示，所以驻波可用 $X(x)T(t)$ 表示。

设 $u(x,t)=X(x)T(t)$ 且 $u(x,t)$ 不恒为零，代入方程和边界条件中得到

$$XT''-a^2X''T=0 \tag{13-5-3}$$

由于 $u(x,t)$ 不恒为零，有 $\dfrac{X''(x)}{X(x)}=\dfrac{T''(t)}{a^2T(t)}=-\lambda$

$$X''(x)+\lambda X(x)=0 \tag{13-5-4}$$

$$T''(t)+\lambda a^2 T(t)=0 \tag{13-5-5}$$

利用边界条件：

$$\begin{cases} X(0)T(t)=0 \\ X(l)T(t)=0 \end{cases} \tag{13-5-6}$$

成立⇔$X(0)=0, X(l)=0$

$$\begin{cases} X'+\lambda X=0 \\ X(0)=0, X(l)=0 \end{cases} \tag{13-5-7}$$

参数 $\lambda$ 成为特征值，函数 $X(x)$ 成为特征函数。

下面分三种情况讨论特征值问题。

(1) $\lambda<0$ 方程的通解为 $X(x)=C_1 e^{\sqrt{-\lambda x}}+C_2 e^{-\sqrt{-\lambda x}}$，由边值条件得

$$\begin{cases} C_1+C_2=0 \\ C_1 e^{\sqrt{-\lambda l}}+C_2 e^{-\sqrt{-\lambda l}}=0 \end{cases}$$

$C_1=C_2=0$，从而 $X(x)=0$，$\lambda<0$ 无意义。

(2) $\lambda=0$ 方程的通解 $X(x)=C_1x+C_2$，同样得到 $X(x)=0$，$\lambda=0$ 无意义。

(3) $\lambda>0$ 时，通解 $X(x)=C_1\cos\sqrt{\lambda}x+C_2\sin\sqrt{\lambda}x$，由边值条件得

$$\begin{cases} C_1=0 \\ C_2\sin\sqrt{\lambda}l=0 \end{cases}$$

得到 $C_2\neq 0$，从而 $\sin\sqrt{\lambda}l=0$，故$\sqrt{\lambda}l=n\pi$，即 $\lambda=\dfrac{n^2\pi^2}{l^2}$，$n=1,2,\cdots$称为本征值，而由于 $X(x)=C_2\sin\dfrac{n\pi}{l}x$，$n=1,2,\cdots$称为本征解或本征函数。本征值是数学物理方程中具有专业意义的概念，其是现代物理学中量子化的数学基础。

再求 $T$：$T''_n(t)+a^2\dfrac{n^2\pi^2}{l^2}T_n(t)=0$

其解为：$T_n(t)=A_n\cos\dfrac{n\pi at}{l}+B_n\sin\dfrac{n\pi at}{l}$

所以 $u_n(x,t)=\left(A_n\cos\dfrac{n\pi at}{l}+B_n\sin\dfrac{n\pi at}{l}\right)\sin\dfrac{n\pi x}{l}$，$n=1,2,\cdots$根据叠加原理可以得到 $u(x,t)=\sum\limits_{n=1}^{\infty}\left(A_n\cos\dfrac{n\pi at}{l}+B_n\sin\dfrac{n\pi at}{l}\right)\sin\dfrac{n\pi x}{l}$。

定解问题的解是傅里叶正弦级数，这是在 $x=0$ 和 $x=1$ 处的第一类齐次边界条件决定的。

$$\begin{cases} A_n=\Phi_n=\dfrac{2}{l}\displaystyle\int_0^l \Phi(\xi)\sin\dfrac{n\pi\xi}{l}d\xi \\ B_n=\dfrac{l}{na\pi}\Psi_n=\dfrac{2}{na\pi}\displaystyle\int_0^l \Psi(\xi)\sin\dfrac{n\pi\xi}{l}d\xi \end{cases}$$

由此得本征值解

$$u_n(x,t)=\left(A_n\cos\frac{na\pi t}{l}+B_n\sin\frac{na\pi t}{l}\right)\sin\frac{n\pi x}{l}$$
$$=N_n\sin(\omega_n t+S_n)\sin\frac{n\pi x}{l} \tag{13-5-8}$$

最后,由叠加原理,得

$$u(x,t)=\sum_{n=1}^{\infty}u_n(x,t) \tag{13-5-9}$$

$u(x,t)$是由无穷多个振幅、频率、初相位各不相同的驻波叠加而成。$n=1$ 的驻波称为基波,$n>1$ 的驻波叫作 $n$ 次谐波。

注意分离变量法适用范围:偏微分方程是线性齐次的,并且边界条件也是齐次的。

其求解的关键步骤:确定本征函数和运用叠加原理。对于不同类型的定解条件有如下的总结,见表 13-5-1。

**表 13-5-1 不同类型的定解条件**

| 左端点 | 右端点 | 本征值 | 本征函数 | 取值范围 |
|---|---|---|---|---|
| 一 | 一 | $\frac{n^2\pi^2}{l^2}$ | $B_n\sin\frac{n\pi}{l}x$ | $n=1,2,\cdots$ |
| 一 | 二 | $\frac{(2n+1)^2\pi^2}{4l^2}$ | $B_n\sin\frac{(2n+1)\pi}{2l}$ | $n=0,1,2,\cdots$ |
| 二 | 二 | $\frac{n^2\pi^2}{l^2}$ | $B_n\cos\frac{n\pi}{l}x$ | $n=0,1,2,\cdots$ |
| 二 | 一 | $\frac{(2n+1)^2\pi^2}{4l^2}$ | $B_n\cos\frac{(2n+1)\pi}{2l}$ | $n=0,1,2,\cdots$ |

### 13.5.4 行波法

基本思想:先求出偏微分方程的通解,然后用边界条件确定特解。

关键步骤:通过变量代换,将波动方程化为便于积分的齐次二阶偏微分方程。

适用范围:无界域内的波动方程等。

**1. 一维情况**

设无界域波动的定解问题:

$$\begin{cases}\dfrac{\partial^2 u}{\partial t^2}=a^2\dfrac{\partial^2 u}{\partial x^2}, & -\infty<x<\infty,t>0\\[2ex] u(x,0)=\Phi(x),\dfrac{\partial u(x,0)}{\partial t}=\Psi(x), & -\infty<x<\infty\end{cases}$$

其解为达朗贝尔公式——行波:

$$u(x,t)=\frac{1}{2}[\varphi(x+at)+\varphi(x-at)]+\frac{1}{2a}\int_{x-at}^{x+at}\psi(\xi)\mathrm{d}\xi \quad (\text{一维的达朗贝尔公式})$$

达朗贝尔解的物理意义:

其一,只有初始位移时,$u(x,t)=\frac{1}{2}[\phi(x+at)+\phi(x-at)]$,$\phi(x-at)$代表以速度 $a$ 沿 $x$ 轴正向传播的波,$\phi(x+at)$代表以速度 $a$ 沿 $x$ 轴负向传播的波。

其二，只有初始速度时，$u(x,t)=\frac{1}{2a}\int_{x-at}^{x+at}\psi(\xi)\mathrm{d}\xi$，假使初始速度在区间上是常数，而在此区间外恒等于0，$u(x,t)=\psi_1(x+at)-\psi_1(x-at)$。

**2. 多维情况**

三维波动方程的初值问题

$$\begin{cases}\frac{\partial^2 u}{\partial t^2}=a^2\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right), \quad t>0, x,y,z\in\mathbf{R}\\ u|_{t=0}=\varphi(x,y,z)\\ \left.\frac{\partial u}{\partial t}\right|_{t=0}=\psi(x,y,z)\end{cases}$$

令 $\Delta u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}$和球对称情形化为球面坐标系：令$\begin{cases}x=r\sin\varphi\cos\theta,\\ y=r\sin\varphi\sin\theta,\\ z=r\cos\varphi,\end{cases}$则

$$\Delta u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right)+\frac{1}{r^2\sin\varphi}\frac{\partial}{\partial\varphi}\left(\sin\varphi\frac{\partial u}{\partial\varphi}\right)+\frac{1}{r^2\sin\varphi}\frac{\partial^2 u}{\partial\theta^2}$$

其解为

$$u(x,y,z,t)=\frac{\partial}{\partial t}(t\bar{\varphi})+t\bar{\psi}=\frac{\partial}{\partial t}\left(\frac{1}{4\pi a^2 t}\iint_{S_{at}^M}\varphi\mathrm{d}S\right)+\frac{1}{4\pi a^2 t}\iint_{S_{at}^M}\psi\mathrm{d}S$$

$$u(x,y,z,t)=\frac{1}{4\pi}\frac{\partial}{\partial t}\left[t\int_0^{2\pi}\int_0^{\pi}\varphi(x+at\sin\phi\cos\theta,y+st\sin\phi\sin\theta,z+at\cos\phi,t)\sin\phi\,\mathrm{d}\phi\,\mathrm{d}\theta\right]$$
$$+\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\psi(x+at\sin\phi\cos\theta,y+st\sin\phi\sin\theta,z+at\cos\phi,t)\sin\phi\,\mathrm{d}\phi\,\mathrm{d}\theta$$

二维波动方程的初值问题

$$\begin{cases}\frac{\partial^2 u}{\partial t^2}=a^2\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right),\\ u|_{t=0}=\varphi(x,y), \qquad t>0, x,y\in\mathbf{R}\\ \left.\frac{\partial u}{\partial t}\right|_{t=0}=\psi(x,y),\end{cases}$$

求解方法：降维法——由高维波动方程的柯西问题的解来求解低维波动方程柯西问题的方法。该方法是由 Hadamard 最早提出的。由于初始数据与第三个变量无关，因此，在 $S_{at}^M$ 上的球面积分可由在圆域：$\sum_{at}^{M}$：$(\xi-x)^2+(\eta-y)^2\leqslant(at)^2$ 上的积分得到。

$$\frac{1}{4\pi a^2 t}\iint_{S_{at}^M}\phi\,\mathrm{d}\omega=\frac{1}{4\pi a}\iint_{S_{at}^M}\frac{\phi}{r}\mathrm{d}\omega=\frac{1}{4\pi a}\iint_{S_{at}^M(+)}\frac{\phi}{r}\mathrm{d}\omega+\frac{1}{4\pi a}\iint_{S_{at}^M(-)}\frac{\phi}{r}\mathrm{d}\omega$$
$$=2\frac{1}{4\pi a}\iint_{\sum_{at}^M}\frac{\phi}{r}\sqrt{a+\zeta'^2_\xi}\,\mathrm{d}\sigma=\frac{1}{2\pi a}\iint_{\sum_{at}^M}\frac{\phi(\xi,\eta)}{\sqrt{(at)^2-(\xi-x)^2-(\eta-y)^2}}\mathrm{d}\xi\mathrm{d}\eta$$

和

$$\frac{1}{4\pi a^2 t}\iint_{S_{at}^M}\phi \mathrm{d}\omega=\frac{1}{2\pi a}\iint_{\sum_{at}^M}\frac{\phi(\xi,\eta)}{\sqrt{(at)^2-(\xi-x)^2-(\eta-y)^2}}\mathrm{d}\xi\mathrm{d}\eta$$

$$=\frac{1}{2\pi a}\int_0^{at}\int_0^{2\pi}\frac{\phi(x+\rho\cos\theta,y+\rho\sin\theta)}{\sqrt{(at)^2-\rho^2}}\rho\mathrm{d}\theta\mathrm{d}\rho$$

可以得到

$$u(x,y,t)=\frac{1}{2\pi a}\frac{\partial}{\partial t}\left(\int_0^{at}\int_0^{2\pi}\frac{\phi(x+\rho\cos\theta,y+\rho\sin\theta)}{\sqrt{(at)^2-\rho^2}}\rho\mathrm{d}\theta\mathrm{d}\rho\right)$$

$$+\frac{1}{2\pi a}\int_0^{at}\int_0^{2\pi}\frac{\psi(x+\rho\cos\theta,y+\rho\sin\theta)}{\sqrt{(at)^2-\rho^2}}\rho\mathrm{d}\theta\mathrm{d}\rho$$

物理意义：三维情况是惠更斯原理(有清晰的前锋和后尾)，

二维情况是波的弥散(有清晰的前锋但无后尾)。

### 13.5.5 积分变换

(1) 傅里叶变换法。

(2) 拉普拉斯方程的格林函数法。

### 13.5.6 格林函数法

格林函数，又称点源影响函数，是数学物理方程中的重要概念，代表一个点源在一定的边界条件和初始条件下所产生的场。而知道了点源的场，可以用叠加的方法计算任意源产生的场。

第一格林公式：$\iint_\Gamma u\,\nabla v\cdot\mathrm{d}S=\iiint_\Omega\nabla\cdot(u\,\nabla v)\mathrm{d}V=\iiint_\Omega u\Delta v\mathrm{d}V+\iiint_\Omega\nabla u\cdot\nabla v\mathrm{d}V$

同理 $\iint_\Gamma v\,\nabla u\cdot\mathrm{d}S=\iiint_\Omega v\Delta u\mathrm{d}V+\iiint_\Omega\nabla u\cdot\nabla v\mathrm{d}V$

第二格林公式：两式相减就得到$\iint_\Gamma\left(u\frac{\partial v}{\partial n}-v\frac{\partial u}{\partial n}\right)\cdot\mathrm{d}S=\iiint_\Omega(u\Delta v-v\Delta u)\mathrm{d}V$

讨论带有一定边界条件的泊松方程的求解问题，泊松方程 $\Delta u=f(r)$，$(r\in T)$。而第一、第二、第三类边界条件可以统一表示为 $\left[\alpha\frac{\partial u}{\partial n}+\beta u\right]_\Gamma=f$。其中 $f$ 是区域边界上给定的函数，$\alpha=0,\beta\neq0$ 为第一类边界条件，$\alpha\neq0,\beta=0$ 为第二类边界条件；$\alpha\neq0,\beta\neq0$ 为第三类边界条件。

三维空间拉普拉斯方程的基本解：

定点是 $M_0(x_0,y_0,z_0)$，动点是 $M(x,y,z)$，有

$$v(M,M_0)=\frac{1}{4\pi r_{M_0M}}=\frac{1}{4\pi\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2}}$$

单位正电荷位于 $M_0$ 处，其电场于 $M$ 点的电位为$\frac{1}{4\pi r_{M_0M}}$。

二维空间拉普拉斯方程的基本解：

定点是 $M_0(x_0,y_0)$，动点是 $M(x,y)$，有

基本解：$v(M,M_0)=\frac{1}{2\pi}\ln\frac{1}{r_{M_0M}}=\frac{1}{2\pi}\ln\frac{1}{\sqrt{(x-x_0)^2+(y-y_0)^2}}$

$$u(M_0)=\frac{1}{4\pi}\oiint_{\Gamma}\left[\frac{1}{r_{MM_0}}\frac{\partial u}{\partial n}-u\frac{\partial}{\partial n}\left(\frac{1}{r_{MM_0}}\right)\right]\mathrm{d}S$$

# 学习资源

## 一、中国近现代著名科学家（物理学）简介

梅贻琦（1889—1962），字月涵，著名教育家，天津人。1909 年考取游美学务处选派的首批留学生赴美留学。1914 年毕业于伍斯特理工学院电机系，获电机工程学士学位。1915 年到清华学校任教，后任教务长，暂代校务。1931 年，梅贻琦出任清华校长，一直服务于清华（包括后来的台湾清华大学），因此被誉为清华的“终身校长”。在他的领导下，清华才得以在十年之间从一所颇有名气但无学术地位的学校一跃而跻身于国内名牌大学之列。他成功地出掌清华并奠定了清华的校格：一是师资人才的严格遴选和延聘，二是推行一种集体领导的制度。“生斯长斯，吾爱吾庐”。梅贻琦用这八个字概述了他与清华的血缘之亲，也表达了他对清华的挚爱。他说：“学校犹水也，师生犹鱼也，其行动犹游泳也，大鱼前导，小鱼尾随，是从游也。”梅贻琦是清华百年历史上四大哲人之一，另外三位是叶企孙、潘光旦、陈寅恪。



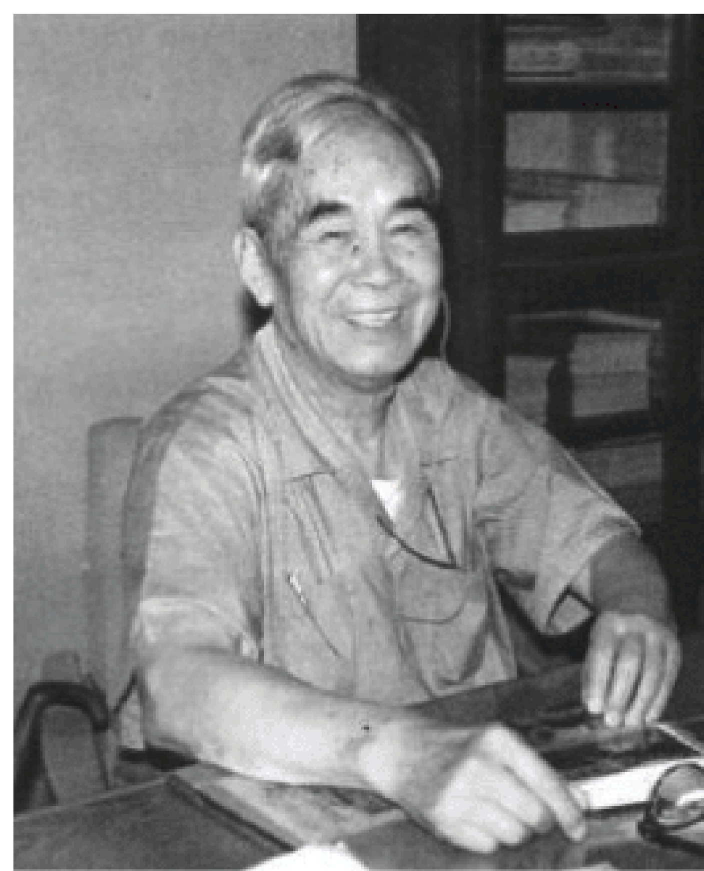
梅贻琦自幼熟读经史，且善背诵。他的同仁回忆说，有一次梅贻琦表示，“假如我们之中有谁背诵任何中国古经传有错漏，我可以接背任何章节。”梅贻琦曾撰文总结自己的留学观，他告诫行将赴美的学生，“诸君在美的这几年，亦正是世界上经受巨大变化的时期，将来有许多组织或要沿革，有许多学说或要变更。我们应保持科学家的态度，不存先见，不存意气，安安静静地去研究，才是正当的办法，才可以免除将来冒险的试验，无谓的牺牲。”

八年抗战期间，梅贻琦仍是清华大学校长，并主持西南联合大学的校务（任联大常委）。西南联大的存在和发展，在我国高等教育史上写下了重要篇章。抗战胜利后，清华大学回到北平，梅贻琦继续担任校长，直到 1948 年 12 月。在他任校长的 17 年里，清华大学得到长足发展，全校设有文、理、工、法、农等 5 个学院 26 个系，在校师生 2400 多人。

梅贻琦于 1955 年由美去台湾，创办新竹清华大学，担任校长直到 1962 年逝世于台北。所以，有人说他是一位“终身校长”。其夫人韩咏华，1977 年由美国回到北京定居，受到邓颖超的热情接待。宴请时，特意由天津狗不理的名厨师做了一桌天津菜。梅贻琦在清华有两句名言，多年来一直为人传诵：一句是：“大学者非谓有大楼之谓也，有大师之谓也。”另一句是：“学生没有坏的，坏学生都是教坏的。”

张文裕（1910—1992）曾用名张少岳，中国科学院院士，中国科学院高能物理研究所名誉所长。福建惠安人，毕业于燕京大学。我国宇宙线研究和高能实验物理的开创人之一。毕生致力于核科学研究和教学，有多项重要发明和发现，学术上最突出的成就是发现 $\mu$ 介原子，开创了奇特原子物理的深入研究。重视实验科学，重视实验基地的建设，为我国高能物

理的发展、北京正负电子对撞机的建成奠定了坚实基础。在核科学领域培养了大批人才。张文裕历任中国科学院近代物理研究所(后相继易名为物理研究所、原子能研究所)研究员、宇宙线研究室主任、副所长,中国科技大学教授及近代物理系主任,十一国联合原子核研究所中国组组长(并领导一个联合研究组),中国科学院高能物理研究所所长兼中国科学院高能物理实验中心筹备处主任,中国物理学会常务理事,中国核学会理事,中国高能物理学会理事长,《中国科学》、《科学通报》主编,中美高能物理合作联合委员会中方主席等职。



## 二、诺贝尔物理学奖简介

**2011 年诺贝尔物理学奖**。通过观测遥远超新星发现宇宙的加速膨胀。获奖者:萨尔·波尔马特(美国)、布莱恩·施密特(美国/澳大利亚)以及亚当·里斯(美国)。评价:三位科学家于 1998 向外公布:宇宙的膨胀速度不是恒定的更不是越来越慢而是不断加速即越来越快。根据牛顿万有引力定律,宇宙大爆炸所产生的冲力在引力的作用下和牵制下,星系天体的退行速度应该渐于趋缓直至稳定平衡,可是这三位科学家的发现却与牛顿万有引力定律相互矛盾,如何解决、诠释这二者之间的矛盾呢?物理学家们认为只有一种可能,那就是宇宙之中存在着一种与引力作用方向相反(反引力作用力),至今人类还没有发现的神秘力量!物理学界把这种与引力作用方向相反、至今人类还未知的、神秘作用力称之为“暗能量”,并且认为,正是这种“暗能量”推动星系天体快速膨胀退行。

**2012 年诺贝尔物理学奖**。突破性的试验方法使得测量和操纵单个量子系统成为可能。获奖者:塞尔日·阿罗什(法国)、戴维·维因兰(美国)。评价:突破性的试验方法使得测量和操纵单个量子系统成为可能。两位获奖者均在量子光学领域研究光与物质间的基本相互作用,这一领域自 20 世纪 80 年代中期以来获得了相当多的成就。他们的突破性的方法,使得这一领域的研究朝着基于量子物理学而建造一种新型超快计算机迈出了第一步。就如传统计算机在 20 世纪的影响那样,或许量子计算机将在 21 世纪以同样根本性的方式改变我们的日常生活。极端精准的时钟在他们研究的推动下应运而生,有望成为未来新型时间标准的基础,而其精准度超越现代铯时钟百倍以上。

**2013 年诺贝尔物理学奖**。描述了粒子物理学的标准模型,其预测的基本粒子被发现。获奖者:弗朗索瓦·恩格勒特(比利时)、彼得·希格斯(英国)。评价:他们描述了粒子物理学的标准模型,并成功预测希格斯玻色子的存在。不久前被欧洲核子研究中心运行的大型强子对撞机通过实验发现。获奖科学家提出的理论解释了粒子如何获得质量,有助于人类了解亚原子粒子质量的起源。瑞典皇家科学院认为,发现相关理论成就巨大,但发现希格斯玻色子并非完成宇宙谜题拼图的最后一块。这一模型仅描述了宇宙可见物,只占到整个宇宙物质的五分之一,更多神秘的未知领域还有待更多科学家去探寻。

**2014 年度诺贝尔物理学奖**。蓝色 LED 技术的应用。获奖者:赤崎勇(日本)、天野浩(日本)以及中村修二(美国,日本公民)。评价:他们发明一种新型高效节能光源方面的贡献,即蓝色发光二极管(LED)。通过蓝色 LED 技术的应用,人类可以使用一种全新的手段产生白色光源。相比旧式的灯具,LED 灯具有更加持久且高效的优点。红色与绿色发光二极管已经伴随我们超过半个世纪,但我们还需要蓝光的到来才能彻底革新整个照明技术领域,因为只有完整的采用红、绿、蓝三原色之后,我们才能产生照亮我们世界的白色光源。但

尽管工业界和学界付出了巨大的努力，但产生蓝色光源的技术挑战仍然持续了超过30年之久。当他们通过半导体产生出蓝色光源时，照明技术革命的大门打开了。白炽灯照亮了整个20世纪，而21世纪将是LED灯的时代。

**2015年诺贝尔物理学奖**。发现中微子振荡，实际证明了中微子是有质量的。获奖者：梶田隆章（日本）和阿瑟·麦克唐纳（加拿大）。评价：他们发现中微子振荡，实际证明了中微子是有质量的。梶田隆章现如今担任东京大学宇宙线研究所的所长、同研究所附属宇宙中微子观测信息融合中心负责人。阿瑟·麦克唐纳是萨德伯里中微子天文台研究所主任，目前是女王大学大学研究主席。

**2016年诺贝尔物理学奖**。物质的拓扑相变。获奖者：普林斯顿和布朗大学的三位科学家戴维·索利斯（美国）和邓肯·霍尔丹（美国）以及米歇尔·克里特里兹（美国）。评价：他们在理论上发现了物质的拓扑相变和拓扑相。戴维·索利斯（David Thouless），拥有英国和美国双重国籍，目前任教于美国华盛顿大学；邓肯·霍尔丹（Duncan Haldane）的研究领域为凝聚态物理学，目前为美国普林斯顿大学教授；迈克尔·科斯特立茨（Michael Kosterlitz）拥有英国和美国双重国籍，研究领域为凝聚态物理学。他们将拓扑学作为工具，他们证明了超导性能够在低温条件下实现，并解释了其背后的原理，以及为何在高温下超导性会消失的原因。

**2017年诺贝尔物理学奖**。发现和并探测到了引力波。获奖者：雷纳·韦斯（美国）、巴里·巴里什（美国）和基普·索恩（美国）。评价：他们为"激光干涉引力波天文台"（LIGO）项目和发现引力波所作的贡献。通过Ligo（激光干涉仪）首次探测到引力波，初始的LIGO设备仅仅是在麻省理工的一个课堂练习，由Rainer Weiss最早提出。引力波最初的发现是在2015年底，2016年初宣布美国自然基金委员会主任France Cordova宣布了第四次引力波的重大发现，而且科学家首次得到了引力波的3D波形图。这项重大发现在诺贝尔奖颁发前宣布意义非同一般。引力波最早是在100多年前由爱因斯坦预测提出，是时空（Spacetime）中的一种波形，产生于宇宙中最奇怪的地方——比如时间最初起源的时候黑洞的边缘。所以找到引力波，就好像是打开了宇宙空间的另一扇窗，能够帮助人类找到更好的方法去观测巨大的、诡异的物质，比如无尽的黑洞以及星球的爆炸。

**2018年诺贝尔物理学奖**。激光物理和高功率光脉冲研究领域的突破性发明。获奖者：阿什金（美国）、莫罗（法国）、斯蒂克兰德（加拿大）。评价：他们在激光物理研究领域的突破性发明！阿什金发现：不管一束激光多么细锐，它的强度总是会由中心向边缘逐渐减弱。因此，激光给粒子的辐射压也会发生变化，将粒子推向光束中间，并把它们维持在中心位置。为了将粒子保持在光束方向上，阿什金又增加了一个高强度透镜来聚焦激光。于是，这些粒子被吸向光强度最大的一点，形成一个光陷阱，后来人们称它为光学镊子。莫罗和斯蒂克兰德的新技术，啁啾脉冲放大技术（CPA），把短激光脉冲放大到史无前例的水平。自从约1960年激光被发明开始，研究者们就试图创造出更强的脉冲。然而，到了20世纪80年代中期，这条道路似乎走到了尽头。对于短脉冲而言，再增加光强，就会把放大材料给烧毁了。取一段短激光脉冲，在时间上拉长，把它放大，再重新压缩成短脉冲。当脉冲拉长之后，峰值功率就会大幅下降，这样就可以在不损坏放大器的前提下放大到更高的倍数。然后把脉冲重新压缩，脉冲的强度就随之急剧上升。

**2019年诺贝尔物理学奖**。对太阳系和宇宙结构及历史的新认识。获奖者：詹姆斯·皮布尔斯（加拿大裔美国）和米歇尔·麦耶（瑞士）和迪迪埃·奎洛兹（瑞士）。评价：该年的

诺贝尔物理学奖表彰了两个研究：其一是对宇宙结构和历史的新认识，其二是在太阳系外首次发现一个绕着类太阳恒星公转的行星。大爆炸模型描述了宇宙在大约140亿年前温度密度极高的初始状态。从那个点开始，宇宙急速膨胀，体积变大，温度变低。大爆炸后仅仅40万年，宇宙变得透明，光线可以在空间中穿行。即使到了今天，大爆炸的余辉仍然以辐射形式存在于我们的周围，宇宙的许多秘密都隐藏在其中。James Peebles使用他的理论工具与计算，解释这些来自宇宙初期的线索，并发现新的物理过程，使人类重新认识宇宙。

## 思考与练习

1. 如何理解高等数学在物理学理论发展过程中的促进作用?
2. 如何理解微积分在力学理论发展中的作用?
3. 如何理解概率统计学在热学理论发展中的作用?
4. 如何理解高等数学在电磁学的发展与建立过程中的作用?
5. 光学研究的理论基础是什么?
6. 你觉得高等数学知识对普通物理学的学习和深入理解有哪些帮助?

# 参 考 文 献


[1] 荣曾齐.物理概念与教材结构[M].北京：科学出版社,1989.
[2] 王沛清.物理概念教学[M].长沙：湖南教育出版社,1985.
[3] 夏宗经.宏观、宇观和微观[M].北京：中国青年出版社,1984.
[4] 教育部师范教育司.二十世纪物理学概观[M].上海：上海科技出版社,1999.
[5] 武际可,隋允康.力学史与方法论论文集[M].北京：中国林业出版社,2003.
[6] 彼德·迈克尔·哈曼.十九世纪物理学概念的发展[M].龚少明,译.上海：复旦大学出版社,2000.
[7] 张三慧. 大学物理学[M].北京：清华大学出版社,2005.
[8] 李忠.物理学概念教学研究[M].北京：清华大学出版社,1998.
[9] 郭奕玲,沈慧君.物理学史[M].北京：清华大学出版社,2005.
[10] 钟霄参.物理学史[M].杭州：浙江教育出版社,1985.
[11] 漆安慎,杜禅英.力学[M].北京：高等教育出版社,1997.
[12] 王楚,李椿,徐安士.热学[M].北京：北京大学出版社,2000.
[13] 大学物理编辑部.力学热学专辑[M].北京：对外贸易出版社,1987.
[14] 贾世楼.信息论理论基础[M].哈尔滨：哈尔滨工业大学出版社,2001.
[15] 顾建中. 普通物理简明教程(力学)[M].北京：人民教育出版社,1977.
[16] 赵凯华,罗薇茵.力学[M].北京：高等教育出版社,2002.
[17] 赵凯华,罗薇茵.热学[M].北京：高等教育出版社,2001.
[18] 曹清喜.热学[M].北京：北京师范大学出版社,1989.
[19] 包科达.热物理学基础[M].北京：高等教育出版社,2001.
[20] 薛增泉. 热力学与统计物理学[M].北京：北京大学出版社,1995.
[21] 马本坤,高尚惠,孙煜. 热力学与统计物理学[M].北京：高等教育出版社,1986.
[22] 汪志诚. 热力学.统计物理学[M]. 北京：高等教育出版社,1980.
[23] 冯玉广,李士. 热力学与统计物理学导论[M]. 北京：中国科学技术出版社,1993.
[24] 顾建中.热学[M].北京：高等教育出版社,1985.
[25] 王鸿谟,龚远芳,游寿星.基础理论物理[M].成都：四川大学出版社,1992.
[26] 惠萍.负绝对温度状态的讨论[J].广东教育学院学报,1998,18(2)73.
[27] 王家庆.关于负绝对温度的几点讨论[J].大学物理,1996,15(8)21-22.
[28] 张可言.负绝对温度系统存在的特征[J].重庆教育学院学报,2000,13(3)：60-62.
[29] 弗·卡约里.物理学史[M].桂林：广西师范大学出版社,2002.
[30] 胡汉平,程文龙.热物理学概论[M].合肥：中国科学技术大学出版社,2006.
[31] 杨冬青,石亚璞.对热力学第二定律的几点认识[J].物理教师,2002,23(6).
[32] 张遥.浅谈热力学第二定律[J].中学生数理化,2005,16：43-45.
[33] 徐汉屏.高中物理中的实际问题——热学部分[J].中学物理教学参考,2002,31(11).
[34] 阳同福.热力学第二定律的等价性证明[J].物理教学探讨,2005,23(7).
[35] 梁绍荣.热学[M].北京：高等教育出版社,2006.
[36] 秦允豪.热学[M].北京：高等教育出版社,1999.
[37] 贾起民,郑永令.电磁学[M].北京：高等教育出版社,1985.
[38] 褚圣麟.原子物理学[M].北京：高等教育出版社,1979.
[39] 张涛光. 物理学方法论[M].济南：山东科学技术出版社,1983.
[40] 程守之,江之永.普通物理学[M].北京：高等教育出版社,2000.

[41] 刘银春.大学物理教程[M].北京：机械工业出版社，2006.
[42] 梁昆淼.数学物理方法[M].北京：高等教育出版社，1998.
[43] 蔡其恕，陈林才，等.长度计量测试基础[M].北京：计量出版社，1986.
[44] 刘毓兰，赵瑞生.长度计量技术[M].北京：机械工业出版社，1988.
[45] 中国计量测试学会科普与教育工作委员会组.新编物理量单位基础[M].北京：中国计量出版社，1995.
[46] 吴百诗.大学物理[M].西安：西安交通大学出版社，2002.
[47] 李绍贵.现代计量浅谈[J].岱宗学刊，1997(4).
[48] 詹继悟.浅谈我国法定计量单位[J].粮食经济与技术，1994(3).
[49] 张士相.国际单位制讲座[M].北京：中国铁道出版社，1983.
[50] 伊里奇·萨依然.国际单位制简介[M].北京：计量出版社，1983.
[51] 小泉袈裟胜.计量单位浅谈[M].王金玉，荫铭，译.北京：计量出版社，1983.
[52] 杨哲，李述.国际单位制[SI]使用知识[M].北京：科学普及出版社，1985.
[53] 俞斯昶.量和单位国家标准名词解释[M].北京：中国计量出版社，1990.
[54] 朱小清，詹云翔.光度测量技术及仪器[M].北京：中国计量出版社，1992.
[55] 曾锋，周小燕.阿佛加德罗常数的测定[J].化学教学，2003(2).
[56] 史蒂芬·霍金.时间简史[M].许明贤，吴忠超，译.长沙：湖南科学技术出版社，2001.
[57] 冯杰，等.国际单位制(SI)基本物理量——时间及其测量方法[J].中山大学学报论丛，2001，21(1).
[58] 姚启钧.光学教程[M].北京：高等教育出版社，2002.
[59] 易明.光学[M].北京：高等教育出版社，1999.
[60] 王楚.光学[M].北京：北京大学出版社，1999.
[61] 冯杰.迈克耳逊干涉仪一般等倾干涉花样的特性分析[J].华南理工大学学报(自然科学版)，1999，27(6).
[62] 朱大成.非线性光学及其发展[J].长春邮电学院学报，1986，2.
[63] 殷兴良.现代光学新分支学科——气动光学[J].中国工程科学，2005，12(7).
[64] E.赫克特，A.赞斯.光学[M].詹达三，秦克成，林福成，译.北京：人民教育出版社，1980.
[65] 竺庆娥.光学中的叠加原理[J].宁夏大学学报，1984(1).
[66] 杨述武.普通物理实验(三、光学部分)[M].北京：高等教育出版社，2000.
[67] 倪光炯，李洪芳.近代物理[M].上海：上海科学技术出版社，1979.
[68] 冯杰.位移电流的热效应探讨[J].华南师范大学学报(自然科学版)，1999(2).
[69] 杨福家.原子核物理[M].上海：复旦大学出版社，1993.
[70] 赵展岳.相对论导引[M].北京：清华大学出版社，2002.
[71] 张功耀.相对论革命[M].长沙：湖南教育出版社，1999.
[72] 斯蒂芬·霍金.万有理论：宇宙的起源与归宿[M].郑亦明，葛凯乐，译.海口：海南出版社，三环出版社，2004.
[73] 霍华德·里奇.《时间简史》导读[M].郑志丰，译.长沙：湖南科学技术出版社，2006.
[74] 阿尔伯特·爱因斯坦.相对论的意义[M].李灏，译.北京：北京大学出版社，2014.
[75] 曾谨言.量子力学(上、下)[M].北京：科学出版社，1982.
[76] 张永德.量子力学[M].北京：科学出版社，2008.
[77] A. M. Portis，H. D. Young.大学物理实验.伯克利物理实验[M].北京：科学出版社，1982.
[78] 李福利.高等激光物理学[M].合肥：中国科技大学出版社，1992.
[79] 张泽瑜，赵均.电动力学[M].北京：清华大学出版社，1987.
[80] 曾谨言.量子力学：卷一[M].北京：科学出版社，2007.
[81] 周世勋.量子力学教程[M].北京：高等教育出版社，1979.
[82] 熊钰庆.量子力学导论[M].广州：广东高等教育出版社，2000.

[83] 王大衍.现代光学与光子学的进展(第二集)[M].天津:天津科学技术出版社,2006.
[84] 刘汉臣,唐远河.波粒二象性对物理概念和方法论发展的影响[J].西北纺织工学院学报,1999,3(13).
[85] 郑庆璋,崔世治.相对论时空观[M].太原:山西科学技术出版社,2005.
[86] 爱因斯坦.狭义与广义相对论浅说[M].北京:北京大学出版社,2006.
[87] 史蒂芬·霍金.果壳中的宇宙[M].吴忠超,译.长沙:湖南科学技术出版社,2001.
[88] 林为民.图说相对论[M].呼和浩特:内蒙古人民出版社,2003.
[89] 福岛肇.相对论的 ABC [M].董湘怀,译.北京:科学出版社,2001.
[90] 孟广达,王润华,等.狭义相对论解决双生子佯谬之不可能[J].大学物理,1997.
[91] 胡毅.双生子佯谬的狭义相对论观点[J].渝西学院学报(自然科学版),2004.
[92] 孙鹏,宋本祥.浅谈双生子佯谬[J].鞍山师范学院学报(综合版),1996.
[93] 谢丽,谭来军.对"双生子佯谬"的辨析[J].广西轻工业,2006,6.
[94] 马和平,马雪峰.挑战哈勃定律[M].大连:大连出版社,2006.
[95] 罗大正.不可视觉物质:暗能量和量子外力[M].成都:四川科学技术出版社,2005.
[96] 当纳·哥尔斯密.爱因斯坦的大错吗[M].徐文秀,绪晓明,译.长春:吉林人民出版社,2001.
[97] 阿尔伯特·爱因斯坦.相对论的意义[M].郝建纲,刘通军,译.上海:上海科技教育出版社,2005.
[98] 武伟轩.宇宙通史[M].北京:台海出版社,2005.
[99] 约阿希姆·布勃拉特.宇宙的秘密[M].王世英,译.保定:河北大学出版社,2003.
[100] 戴念祖,周嘉华.原理——时代的巨著[M].重庆:西南交通大学出版社,1988.
[101] 梁淑娟,苏曾燧.微积分在力学中的应用[M].北京:人民教育出版社,1980.
[102] 傅海伦.中外数学史概论[M].北京:科学出版社,2007.
[103] 樊映川.高等数学(上、下册)[M].北京:人民教育出版社,1979.
[104] F.瑞夫.统计物理学《伯克利物理教程》第五卷[M].北京:科学出版社,1979.
[105] 泡利.泡利物理学讲义[M].宛之方,译.北京:人民教育出版社,1981.

# 后　　记

本次改版的《普通物理专题研究教程(第 2 版)》,具有以下新的特色。其一,总结了近二十年来在讲授、讨论、作业要求和考试成绩评定等“教”的方面的教学经验；其二,收集了两个专业方向硕士研究生在普通物理学理论知识提高方面的主要疑难杂症,分析了学生对普通物理学知识链接的支离破碎、理解上的缺失和歧义所在,比如,对牛顿运动定律逻辑结构的错误认知、电磁学中关于“电场”和“磁场”的物理意义及其数学表述的模糊；对牛顿第一定律的片面理解、基本上不会应用非惯性系思考力学问题；不能够领会叠加原理在力学、光学和电磁学中的物理意义；特别是普遍地不会运用高等数学思考物理问题等“学”的方面的信息；其三,梳理了普通物理学的逻辑结构,从哲学和逻辑学的角度,增加了第 1 章,将物理学的基本概念范畴分为四对概念范畴:“实物与场”“粒子与波”“时间与空间”“运动与静止”；其四,对接于基础教育的课程改革要求,补充了对应于基础教育物理新课程的现代物理学相关内容,比如,相对论、量子论和非线性物理学的基础原理；其五,基于高等数学一直是当前我国大学理科学生的薄弱环节的现实情况,而物理学的形式表征和物理意义必须借助于高等数学的特质要求,增加了第 13 章,物理学中的数学方法,等等。同时,本版突显普通物理的哲学、逻辑学和社会学的科学价值、技术应用价值和教育功能,比如,普通物理的热学,其理论原理是现代交通技术、空调技术的基础,其应用广泛度仅次于微电子技术的电磁学；热学中最具有哲学意义的是热力学第二定律,最具有社会学价值的是“熵”“熵增原理”“信息熵”等,最具有普遍意义是热力学第一定律等。因此,“物理课程与教学论暨学科教学(物理)”硕士研究生、物理教育和科学教育本科生通过该课程的学习,无疑是其对物理学原理、体系和研究方法的深入理解和有效的升华,为其从事物理教学教育及其研究等工作提供了重塑的必要基础。

本次改版除了要继续感谢我的研究生同学们为本书收集、整理素材和参与研究付出的辛勤劳动之外,还要特别感谢复旦大学的冯吴亮博士、内蒙古师范大学的孙咏萍副教授和四川师范大学的帅晓红教授的贡献。冯吴亮博士根据自身学研物理学的体会,积极参与构建第一章的物理学基本概念体系的四对概念范畴,承担了该章的大部分内容的撰写和独立撰写了第 13 章物理学中的数学方法,整理了“诺贝尔物理学奖简介”等许多有价值的素材；孙咏萍副教授独立撰写了第 12 章非线性物理世界,收集整合了各章的“学习资源”；帅晓红教授对这次改版的全部文稿进行了细致的阅读,并提出许多宝贵的修改意见。

在此付梓之际,笔者还要特别感谢上海师范大学数理学院和全国物理教育的领导和同事的关心和支持！最后,特别感谢清华大学出版社的支持和帮助！

由于编著者水平所限,书中错漏之处再所难免,恳请广大读者批评指正。

冯　杰

2003 年 12 月 23 日初稿于广州

2010 年 3 月修订于上海

2019 年 10 月再修订补充于上海